A Neural Network Approach to MR and CT Image Understanding

R. POLI*, D. CARAMELLA*, M. RUCCI*, G. VALLI*

*Dept. of Electronic Engineering, University of Florence, Italy
*Dept. of Radiology, University of Pisa, Italy

Summary

The problems encountered in the development of automatic systems for MR and CT image analysis are firstly recalled. An approach based on the integration of artificial neural networks and computer vision techniques which seems to be capable of overcoming such problems is then presented. According to this approach we have developed a system for the construction of 3-D descriptions of the organs imaged in sequences of MR or CT slices. The architecture and the preliminary results of this system are also described.

Introduction

Magnetic Resonance (MR) and Computed Tomography (CT) are among the most applied digital imaging techniques. They produce a set of tomograms representing sections of the body on a sequence of parallel planes. According to the definition of Udupa [1], the set of 2-D digital images is termed discrete 3-D scene and it is a 3-D array of items, called voxels, whose value represents the local measure of a property of a tissue.

Usually discrete 3-D scenes are analyzed in two dimensions, by sequentially observing the set of tomograms on a screen. In this way, only part of the three-dimensional information contained in the discrete 3-D scene can be recovered by the observer and, furthermore, in qualitative terms only. This is the basic reason why in the last few years some computer vision systems have been developed to extract and quantitatively process 3-D information from discrete 3-D scenes [2,3]. The first aim of these systems is to build a 3-D description of the scene that can be later used for automatic quantitative measurements and for displaying purposes. However, the difficulties of the operations involved in the fulfillment of such a goal have given rise to a trade-off between quality of results and operator intervention. Such systems are not widespread, mainly due to these automation limits. The operations that most often require human interventions are the segmentation of the scene and the interpolation among sections. Especially the segmentation step requires particular care, since it is heavily hampered by the image noise and by the variability of anatomical structures.

The encouraging results obtained in the past few years by artificial neural networks [4,5] seem to indicate them as suitable tools for overcoming the abovementioned drawbacks.
Network Approach

The construction of a description of the imaged organs implies, as a first step, the segmentation of the discrete 3-D scene. The purpose of segmentation is to split the scene into meaningful regions that can be later used for building a final 3-D description. Unfortunately, as stated by Marr [6], in spite of many efforts, the results achieved in segmentation are usually disappointing. This is mainly due to two factors. Firstly there is a theoretic reason which consists of the lack of any analytical formulation of what a meaningful segmentation of a scene means. As a result, up to now, optimal segmentation algorithms have not been devised. The second factor is that almost all the computer vision algorithms so far developed for such a task have been designed to deal with just one type of information at a time (for example regions, edges or texture). This implies that all the other different types of information which are present in the scene are not used in the segmentation process. On the contrary, they act as noise [7].

The first problem can be solved by making use of neural networks and by arranging them in a suitable architecture. Obviously, the lack of formulation does not sustain if the rough shape and other characteristics of the objects to be extracted are known a priori. In this case, for each object in the scene, the segmentation purpose is perfectly clear. Thus, an optimal segmenter can be obtained by using a set of suitably trained neural networks for the extraction of each object of interest.

Furthermore, neural networks are extremely well suited to deal with different types of information at once, so that the second limit can be overcome, too. For instance, after the learning phase, a neural segmenter can behave like a mixture of an edge-detector, a texture-detector and a thresholding device.

Thanks to this peculiarity, neural nets can also combine a priori knowledge about the typical shape of the structures to be extracted with the features actually detected in the images. This top-down action can be extremely useful in the presence of noisy, incomplete or inconsistent data.

The prototypical shape of an object is usually expressed in a reference system centered somewhere on the object itself. Thus, the performances of a network devoted to organ extraction, which utilizes this type of knowledge, are very sensitive to translation. This problem can be overcome if the organs are always presented to the network in the same relative position. It can be obtained by preceding each neural segmenter by an attention-focusing module which sends a sub-image centered on the organ to the subsequent network.

On the basis of these observations, we developed a system which has the structure described in the following section.
System Architecture

The system consists of two parts: a neural section which performs the segmentation of the scene and a 3-D rendering module that refines the description obtained by the first section and displays the organs contained in the scene. All the networks of the system here described are feed-forward nets and they are trained with the backpropagation algorithm. According to what we pointed out in previous section, the neural part of the system includes a vision sub-system for each anatomical structure of interest. Each sub-system is made up of two cascaded modules: the Attention-Focusing Module (AFM) and the Region Finding Module (RFM), so that the system assumes the structure shown in Fig. 1.

![System Architecture Diagram](image)

Figure 1. System architecture.

The tomograms of the discrete 3-D scene are applied sequentially to the inputs of the system. For each image the AFMs give two outputs: a set of regions containing the organs of interest (ROIs), and the position these regions have in the original tomogram. The ROIs are obtained in two steps (see Fig. 2): at first a pyramidization is performed so as to lower the resolution of the tomogram; then a set of appropriately trained neural networks is applied to the low-resolution image.

The pyramidization process gives 16x16 images from 256x256 pixel wide tomograms. In spite of this reduction, an averaging stage guarantees that the low-resolution images still contain the basic anatomical structures.

In each AFM, the pixels of the 16x16 image are the inputs of a neural network which has four fully-connected layers consisting of 256 (16x16), 64, 32 and 32 units, respectively. The 32 output units can be thought of as two sets of 16 (one set for each reference axis) which codify the ROI position. Obviously, the ROI dimensions depend on the selected organ, i.e. on the vision sub-system.
The ROIs are sent to the corresponding RFMs which determine which pixels belong to the considered organs. As shown in Fig. 3, each RFM consists of a five-layer neural network whose inputs are the pixels contained in a squared area (retina) and the position this retina has on the ROI. The RFM can be seen as a non-linear space-varying filter applied to the ROI. The position of the retina is codified by the coordinates of its central pixel on the ROI; the activation of the single output unit specifies if this pixel belongs to the considered organ.

The RFM is structured so as to preprocess independently the two different kinds of information (the values of the pixels and the position of the retina). In fact it can be thought of as two different three-layer fully-connected sub-nets which join in a common fourth layer. The number of units that make up the sub-network which operates on the position...
of the retina may change for each sub-system (depending on the dimension of the considered ROI). On the contrary, the other sub-net is always made up of 81 (9x9) input units, 10 units in the first hidden layer and 10 in the second one. The common fourth layer includes 3 units connected to the aforementioned sub-nets and to the single output unit.

By using a suitable threshold the filtered ROIs assume binary values, which represent the final segmentation of the discrete 3-D scene.

The description of the discrete 3-D scene just built includes (for each organ of interest):

i. A generalized axis (provided by the AFM),

ii. For each slice of the scene, the pixels representing the considered organ (supplied by the RFM).

In spite of its simplicity, this first representation can be used for many tasks such as the measurement of diagnostically meaningful parameters. However, for display purposes, it needs to be improved.

A 3-D rendering module refines this raw representation with an interpolation process which adds, as a third term, a set of contours which fill the gap between adjacent slices. At present, the implemented technique is the dynamic-elastic interpolation algorithm [8]. After the interpolation phase, the selected organs can be shown on the computer display with a conventional shading method (ray-tracing or depth-buffer) for interactive analysis. The image in Fig. 4 is an example of 3-D visualization of a thoracic vertebra, obtained from three adjacent CT images.

Figure 4. Typical system output.
As seen in the previous section, the first aim of the system is the construction of a 3-D description of the shape of the imaged anatomical structures. The way this description is obtained implies the recognition (labeling) of the organs, too. Thus, in some respects, the system can be said to "understand" the scene. At this point, the visualization of the organs is just one of the possible operations. The others include 3-D measurements and further processing.

Artificial neural networks have allowed the integration of different information sources, which in turn have allowed a complete automation of the system. The computation load required for the building of the scene description is not heavy and the parallel organization of the neural section of the system can be further exploited to improve the overall speed. As an example, on a 20 MHz 80386 CPU equipped with mathematical coprocessor, the computation time is about 90 seconds per slice.

Acknowledgments

This research has been supported by grants of Italian M.U.R.S.T.

References


