A neural network to visually understand and autonomously navigate unknown environments

S. Taraglio[1], A. Zanela[1], M. Salerno[2], F. Sargeni[2], V. Bonaiuto[2]

[1] *ENEA - C.R. Casaccia, Via Anguillarese, 301- 00060 Rome, Italy*
[2] *Department of Electronic Engineering ,University of Rome "Tor Vergata", Via di Tor Vergata, 110 - 00133 Rome, Italy*

Abstract. A complex sensor based control system is presented. The sensor used is a pair of TV cameras providing a stereogram for a stereo vision system based on a cellular neural network (CNN). The 3D information retrieved are used by a simple path planning algorithm and actuated on two different robots. Results of experiments are reported. The usability of the CNN paradigm in robotics applications is demonstrated.

1 Introduction

Autonomous navigation requires three-dimensional information about the environment, in order to avoid collisions with moving objects or with the obstacles of the architectural or natural background [1]. One of the possible methods that can be employed to feel the environment in autonomous robotics, is artificial vision. This is extremely interesting since man uses almost exclusively the eyes to understand his surroundings. Unfortunately, while man possesses an extremely efficient image processing/image understanding system, computers have not reached a comparable efficiency and capability. In any case, whatever sensorial input has been chosen, the unavoidable constraint in any kind of sensory-based navigation system is represented by time. The need for a real time processing of the huge amount of input data involved, has opened the way for different classes of algorithms.

A stereo vision algorithm correlates the conjugate points on the two images composing an input stereogram in order to recover the three dimensional information about the environment. Recently a neural implementation of the algorithm has been introduced by some of the authors [2]. It is based on the spontaneous energy relaxation process of a cellular neural network (CNN) together with the formulation of the stereo matching problem as a variational one, with constraints.

The CNN [3] are analogue circuits able to process signals in real time. While evolving they minimise a scalar function that represents the internal energy of the network. The high parallel analogue processing rate and convergence speed of this class of neural architecture makes the CNN paradigm really appealing in such problems where real-time replies to external stimuli are required. A basic issue is represented by the possibility for this neural architecture to migrate towards an actual hardware implementation [4].

In this work a pioneer robotics application using this neural architecture is presented. The three dimensional information obtained from the neural stereo vision system is used to reconstruct the ground map of the environment. The mapping process is based on the occupancy grids approach [5]. Through this approach it is possible to obtain an integrated description of the robot's surroundings, fusing separate local sensor maps. The subsequent planning of the path, that the robot must follow, is performed transforming the cell representation into a graph. The minimum cost path between the initial and goal nodes is computed using an algorithm similar to the A^* one [6].

This work is intended as a practical demonstrator of the usability of the CNN paradigm in real applications. Thus its focus is on the cellular neural network implementation, for the planning step it has been employed a very standard approach.

2 CNN approach to stereo vision and volume reconstruction

From a pair of stereo images it is possible to retrieve depth information, since a given point in the space is seen from slightly different points of view in the two images. The process of matching the conjugate points can be classified into two main approaches: feature-based or area-based. In the first, given features of the images (e.g. contours or edges) are matched, while in the second the correlation of neighbourhoods of pixels is performed. Naturally the first approach produces sparse disparity maps, while the second outputs dense maps.

There also has been considerable research towards the development of algorithms capable to yield dense disparity maps through the simultaneous solution of the correspondence problem for all the image pixels. These algorithms try to compute the disparity function via the minimisation on the whole image of an energy functional representing the problem. As it is well known, the stereo matching problem is inherently an ill posed one. But the regularisation of the problem is possible through the use of a variational approach under some restrictive hypotheses such as the absence of occluded pixels and a smoothing term in the energy function in order to produce a small disparity gradient [7]. The common characteristic of all these variational approaches is being computationally very intensive. Nevertheless a neural network implementation, being based on a different computational philosophy, may yield a very interesting timing performance.

The Cellular Neural Networks feature the double aspect of both the neural networks and the cellular automata. The connectivity of the single cell and the dependency on the activation values of the neighbours recalls the basic feature of the cellular automata paradigm. At the same time it is a large scale analogue circuit composed of a massive aggregate of connected circuit clones; in this it resembles the neural network approach. The stability of a CNN is guaranteed by the existence of a Lyapunov function. This function can be interpreted as the internal energy of the CNN. The *program* in such devices, similarly to the standard neural networks, is stored in a connectivity pattern typically limited to the first neighbours.

In [2] the stereo vision problem has been handled through a variational approach. The CNN can be used as an optimising tool in order to solve a problem expressed in a variational form. Through the comparison of the energy expression of the stereo matching problem, as coded via a CNN, and its internal energy function, the Lyapunov function, it is possible to derive the connection templates that specialise a CNN to the desired application.

Once that the distance is known, through very simple geometry it is possible to reconstruct all the co-ordinates of the visible pixels of the image and navigate. The simplest way to represent this information on the environment is to slice the three dimensional representation of the data in order to obtain a ground map of the surroundings, as seen through the TV cameras, see Figure 1.

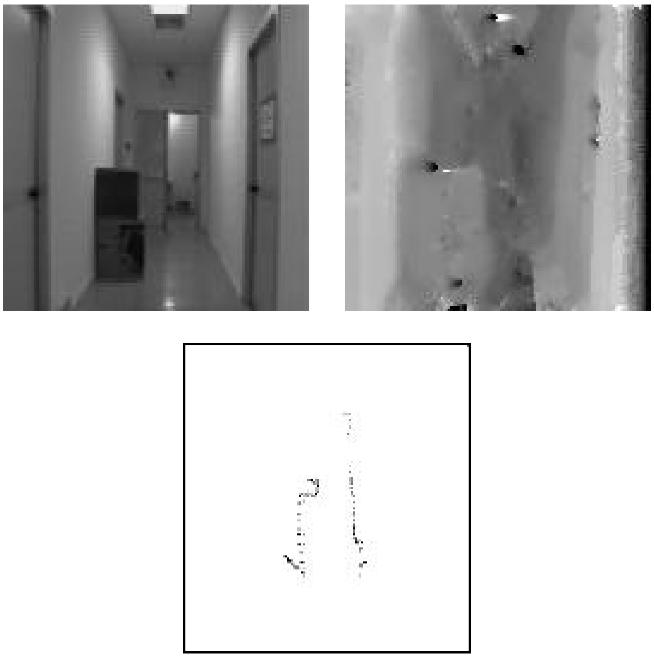


Figure 1. The input image, the disparity map and the obtained ground map via the Stereo-CNN.

Naturally there can be different approaches to retrieve the ground map from video images (i.e. [8]), in this work, though, the primary focus has been the demonstration of the usability of the CNN for autonomous navigation, this is also the reason for the utilisation of the full reconstruction of the environment, being this the natural output of the Stereo-CNN algorithm and thus at no extra computational cost. Besides the possibility to utilise a full three dimensional reconstruction of the environment can be extremely useful in other robotics issues, for example self localisation or shape recognition, different from the actual navigation, but important for the robot autonomy.

3 Hardware implementation

In [9] the authors presented some results about the preliminary tests of the Stereo-CNN algorithm implemented on a general-purpose CNN hardware [4]. The very promising results of these tests suggested to design and manufacture a new CNN board better suited for that particular application.

The design of the board is still based on the original chip [10] as the CNN analogue processing core. Four of these neurochips have been connected together to implement a network of 6x24 cells. Furthermore, the board has been interfaced using the PCI interface, to place it directly on the Pentium based PC, which is on board of the robot, reducing any delay related to the data transfer. The board has been equipped with a 16-bit micro-controller to handle all the on-board operations (e.g. data upload, download, start network, etc.). A PCI target controller assures the interface signal handling. A SRAM of 512x16 Kbytes has been placed on the board to store the images. In addition, some ADC's will manage the proper conversion of the input/output analogue state voltages allowing the acquisition of the output grey scale images. The host PC will acquire the images and store them directly on the SRAM via the PCI interface. The controller reads the data on the SRAM and transforms them to feed the analogue CNN processing core. Finally, it gathers the results from the CNN analogue processing core computing the output disparity map.

The board is presently being assembled and the expected processing rate for a 48x48 pixels grey scale image will be about 10 frame/s.

Further evolution of the hardware is represented by the ongoing design of a new CNN chip explicitly tuned for the Stereo-CNN algorithm. There will be a higher degree of integration and it will be possible to place on the same chip a larger number of Stereo-CNN cells. At the same time there will be an increase in time performances. The foreseen performances can be estimated in circa five times increase in the image dimension and ten times the present velocity.

It is important to stress that these custom chips are low cost ones and that the core of the chip is represented by the actual implementation of an analogue neural network. Some people may oppose that the ever increasing power of processors may offer low cost, real time performances to algorithms that are still computationally too intensive, and that the realisation of custom hardware may be a technological *cul-de-sac*. This point of view may indeed be true, but it is our belief that the search for different paths and the comparison of different approaches in the research has been, is, and it will always be stimulating and fruitful. Besides the application here described represents an important step to demonstrate the usability of the CNN paradigm in real systems.

4 The planner and navigator subsystem

Occupancy grids are a well known and reliable method to *fuse* multiple sensor readings into a global map of the environment. In this work the only sensor used is a stereo vision system which is made operate during the motion of a robotised platform. Thus the fusion is performed both in space and in time. The data to be fused are ground maps, obtained as above explained, in order to obtain a more reliable representation of the environment in which the robot has to move. The process of fusion is carried out through an additional parameter of the generic map pixel, the occupancy reliability. This parameter possesses an initial value of 0.5 and is constantly updated, to keep track of the added sensor readings. The ground map is divided into a set of squared cells $C(i,j)$ of the physical dimensions of the robot base. For each cell a state is defined which possesses a value depending on its state of occupation, see Figure 2. If the cell in the grid is considered as occupied, then it will be excluded in the free-path calculation.

The computation of the path is performed via the generation of a graph, where each node represents a cell of the occupancy grid and the arcs connect the cells in a nearest neighbour structure. If a given cell is occupied, the relative node is ruled out in the actual computation. Path planning avoiding obstacles is performed using a search algorithm in this graph similar to the A^* algorithm [6].

5 Experimental results

The results here presented are relative to experiments performed in indoor and outdoor environments, i.e. partially structured ones in the sense that there are preferential straight lines and planar surfaces. We here present the navigation in a corridor performed by the robotised platform “Tersy”, based on a commercial system B21 of the Real World Interface [11], in the framework of the TERSYCORE project. The preliminary results in the framework of the PRASSI project for outdoor robotics are also presented. The overall software organisation of the robot is realised in a client-server architecture and the main control loop of this application is composed of the following three steps. Grabbing of the stereogram and neural processing. Three dimensional reconstruction, ground map and planning. Actual move.

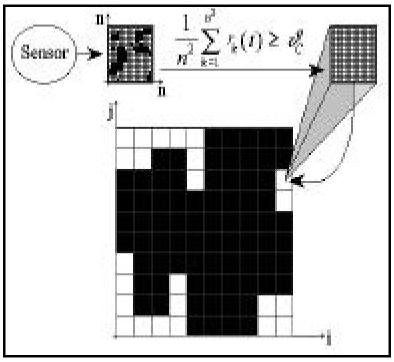
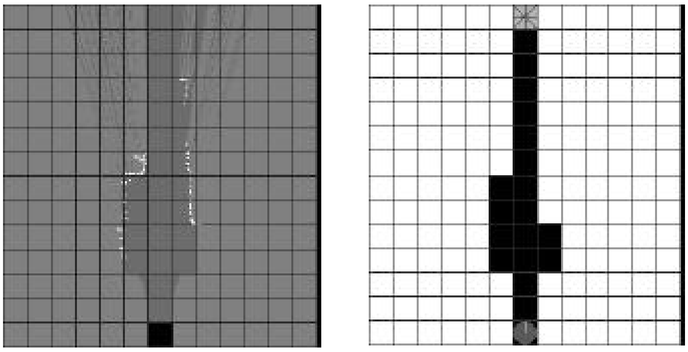


Figure 2. The computation of the state of the generic cell *C(i,j)*.



(a) (b)

Figure 3. The occupancy maps obtained with the data of Figure 1.

As above said, presently the CNN is simulated on one of the on board computers, with a timing of about 15 seconds for a single step of the control loop. Using the above presented hardware CNN board a real time performance is expected.

In Figure 3.a is presented the map obtained from the ground map in Figure 1. In this image, corresponding to the initial sensor view, four classes of points can be found, represented with different grey levels (a darker level means a point with a higher probability of being free), black pixels represent free locations (here the location of the robot, i.e. the locations that robot has already visited). The detected obstacles are marked using a lighter grey level. Points for whom no data are available (e.g. points out of the field of view or occluded) are left to a neutral grey, or to the value relative to the previous step.

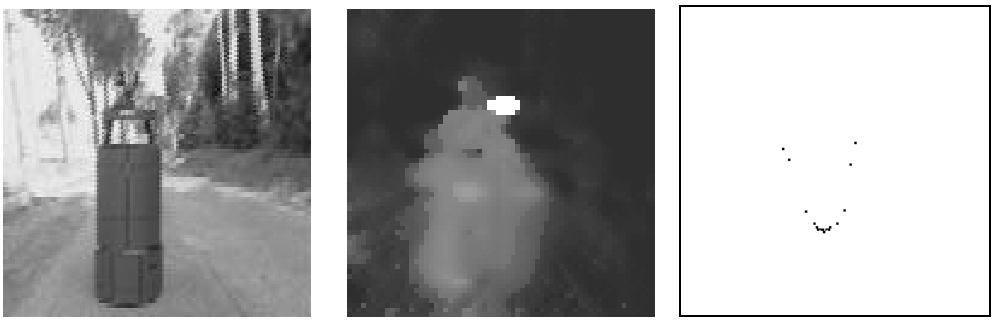


Figure 4. Outdoors experiment. Input image, disparity map, ground map.

In Figure 3.b is shown the obtained occupancy map that will be used to compute a free-collision path from the current robot position to the goal. In the Figure the goal point is relative to an initial mission task definable as *go straight forward as long as possible*. In indoor experiments the robot has been asked to follow a corridor where some obstacles can be found. These are opportunely placed in order to impair the possibility of following the minimum energy path. Tersy has proved able to detect and avoid the obstacles.

In the corridor experiment (Figure 3) the average width of the corridor is retrieved as 182 cm, while the actual measure is 190 cm. The obstacle has been measured at 363 cm with a width of 48 cm, the actual measure is 345 cm with a width of 45 cm. The average percentage error is of about 4%. In the preliminary outdoors experiments the error has been larger, of the order of 7%, this is probably due to the noisier input data, see Figure 4.

6 Conclusions

A complex sensor based control system has been presented. The sensor used is a pair of TV cameras providing a stereogram for a stereo vision system based on a cellular neural network. The information thus extracted is used to create a representation of the environment in which the robot moves. The obstacles and the architectural structure of the space are thus detected and a free path is computed and actuated.

The presented results have been obtained with a digitally simulated neural network, but, the current customisation of the available hardware CNN systems will allow a real time version of the system. The usability of the CNN paradigm in robotics applications has thus been demonstrated.

Acknowledgements

The authors are indebted with the helpful comments of the referees.

This work has been partly funded by the Italian Ministry for the University and the Scientific and Technological Research in the framework of the PRASSI project of the "Calcolo Parallelo con Applicazioni alla Robotica" (High Performance Computing for Robotics Application) initiative.

References

[1] M.A. Salichs, A. Halme, editors. *Proc of the 3rd IFAC Symposium on Intelligent Autonomous Vehicles*, Madrid, Spain, 1998.

[2] S. Taraglio, A. Zanela. A practical use of cellular neural networks: the stereo-vision problem as an optimisation. *Machine Vision and Applications*, **11**:242-251, 2000.

[3] L.O. Chua, Y. Yang. Cellular Neural Network: Theory. *IEEE Transaction on circuits and systems*, **35**:1257-1272, 1988.

[4] M. Salerno, F. Sargeni, V. Bonaiuto. A 720 cells interconnection-oriented system for cellular neural networks. *Proc of IEEE Int Conference on Circuits and Systems (ISCAS 97)*, Special Session "Learning on Silicon", pp. 681-684, Hong Kong, 1997.

[5] A. Elfes. Using occupancy grids for mobile robot perception and navigation. *Computer*, pp. 46-57, June 1989.

[6] P.E. Hart, N.J. Nilsson, B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transaction on systems, science and cybernetics*, **2**:100-107, 1968.

[7] T. Poggio, V. Torre, C. Koch. Computational vision and regularization theory. *Nature*, **317**:314-319, 1985.

[8] H.A. Mallot, H.H. Bülthoff, J.J. Little, S. Bohrer. Inverse perspective mapping simplifies optical flow computation and obstacle detection. *Biological Cybernetics* **64**:177-185, 1991.

[9] S. Taraglio, A. Zanela, M. Salerno, F. Sargeni, V. Bonaiuto. A CNN based vision hardware system for autonomous robot navigation. *Proc Fifth IEEE International Workshop on Cellular Neural Networks and their Applications*, CNNA-98, pp. 181-185, London, 1998.

[10] M.Salerno, F.Sargeni, V.Bonaiuto. A 6x6 cells interconnection-oriented programmable chip for CNN. *Analog Integrated Circuits and Signal Processing*, Kluwer Academic Publisher, **15**(3):239-250, March 1998.

[11] http://www.rwii.com