ELMAN NEURAL NETWORKS AND TIME INTEGRATION FOR OBJECT RECOGNITION

STEFANO ZAPPACOSTA, GIANLUCA BALDASSARRE, AND STEFANO NOLFI

ABSTRACT. We consider a system based on an Elman network for a categorization task. Four objects are investigated by an automa walking around in circles. The shapes are derived from four version of a cross: square, thick cross, critical cross and thin cross. Therefore, the input of the system is represented by the distance-wave relieved by the sensor at each step. We let several parameters vary: starting point and speed of the automa walk, radius of the circle and size of the shape. The system is trained using a back-propagation algorithm. We describe the complete setup of the parameters and noises, which the automa will have to face for the prediction/categorization task.

INTRODUCTION

In this paper we consider a model of an automa endowed with a distance sensor, which is able to walk around objects. On one hand, we are interested in the the capacity of that automa to categorize the object within a reasonable class of objects, and on the other hand robustness of such a capacity with respect to variations of the parameters involved in the entire setting, and with respect to noise parameters (see [10], [14], [9] and [13]).

The model is reduced in our case to facing sequences of distance predictions, which correspond to special object contours. That task play a key role more in general in the autonomous robotics field, and namely in artificial active vision. Those distance sequences may be processed by an artificial neural network, which tries to reproduce biological neuron-like system. There are many different types of such networks; nevertheless we shall focus our attention on **Elman** neural networks (see [6], [3] and [2]).

The Elman neural network we consider has the following architecture. It is a feed-forward network with two layers (see Figure 1 below). The activation functions for the hidden and the output layer are logistic functions

$$\Phi_\sigma(x) = \frac{1}{1+e^{\sigma x}} \ , \tag{1}$$

where the parameter σ will be set to 1. The learning rule we use is the error back-propagation algorithm (see [11]). Moreover, the learning rate λ will be chosen later on.

This type of network differs from conventional ones in that the input layer has a recurrent connection with the hidden one (denoted by the dashed lines in the Figure 1). Therefore, at each time step the output values of the hidden units are copied to the input ones, which store them and use them for the next time step. This process allows the network to memorize some information from the past, in such a way to better detect periodicity of the patterns (perception task).

Since our network receives the distance from the pattern as input step by step, we are dealing with a passive perception (see [7], [16], [8], [5]). We neglect for the moment interaction with the patterns and active perception (see [12], [15], [1]). Anyway the predition is the main task we shall be involved with (see [4] and [17]).

The object patterns we are interested in are cross-shaped objects with variable lengths in the arms and variable ratio between the longer and the shorter arm. We shall investigate which is the best class of such patterns to be categorized. Our experiments are split in two different periods. The first is the so-called *learning phase*, devoted to train the net and to update the weights. The second is the *testing phase*, devoted to test the answers of the net with the resulting fixed weights.

Date: July 3, 2006.
2000 *Mathematics Subject Classification.* Primary: 92B20; Secondary: 68T10.
Key words and phrases. Artificial Neural Networks, Elman Neural Networks, Object Recognition, Time Integration.
The authors are grateful to the Institute of Cognitive Sciences and Technologies (Rome, Italy).

Section 1 is devoted to the complete analysis of the architecture of the Elman neural network (and the related parameters) and the description of the implemented methods. A motivation for the choice of te patterns is given as well.

1. Methods

The Elman neural network we consider has the following architecture. It is a feed-forward network with three layers: an input layer, a hidden layer, and an output layer (see Figure 1 below). This type differs from conventional ones in that the input layer has a recurrent connection with the hidden one (denoted by the dotted lines in the Figure). Therefore, at each time step the output values of the hidden units are copied to the input ones, which store them and use them for the next time step. This process allows the network to memorize some information from the past, in such a way to better detect periodicity of the patterns.

We choose the following architecture. The main characteristic is the copy of the N_h units from the Hidden Layer to the Input Layer, which has $N_h + 1$ units, since the first unit is given by the (normalized) distance measured by the sensor. Moreover we notice that the Output Layer has $1 + N_o$ units, where N_o denotes the total number of patterns (to be categorized), whereas the first unit is devoted to the prediction of the next distance step.

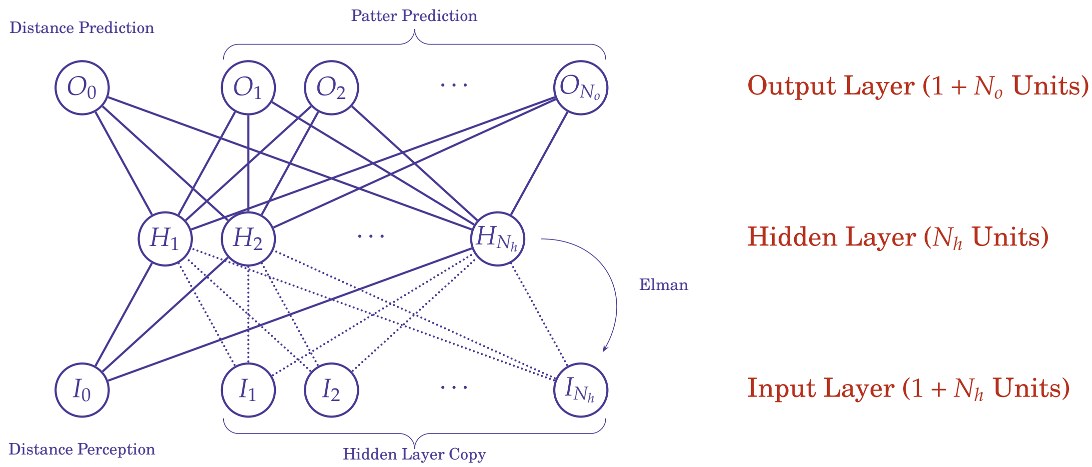


Figure 1: The architecture of the Elman Neural Network

The setting for the automa and the pattern is sketched in an xy-plane in Figure 2. Namely, the path of the automa and the pattern have the following four parameters:

Steps: The total number of steps the automa needs to cover a single lap around the pattern. We shall denote it by *#Steps*. Because of symmetry we shall consider multiples of four, therefore we shall restrict to the reasonable values

$$\#Steps \in \{36, 40, 44, \ldots, 148\} . \tag{2a}$$

The mean value for this parameter is 92 (that is 23 steps for each quadrant).

Starting Point: The angle the automa starts its circle around the pattern. The single angle-step of the automa will depend on the total number of steps for lap, that is $\frac{2\pi}{\#Steps}$. If we let $\alpha(n)$ denote the angle at the n-th step, we have that

$$\alpha(n) \in \left\{0, \frac{2\pi}{\#Steps}, \frac{4\pi}{\#Steps}, \ldots, (\#Steps - 1)\frac{2\pi}{\#Steps}\right\} , \qquad \forall n = 0, \ldots, \#Steps . \tag{2b}$$

The parameter $\alpha(0)$ may be either fixed or chosen randomly.

Radius: The radius of the circle followed by the automa around the pattern, denoted by R. Yo mimic range limitations related to the sensor of real robots we assume the following restrictions

$$R \in [2, 3] . \tag{2c}$$

Notice that the mean value for the radius is 2.5

Size: The size of the pattern, denoted by S. It represents half of the longest arm of the cross pattern. We also let T denotes half of the length of the shortest arm of the cross (see the Figure 2). We shall assume some restriction of the size, as well

$$S \in [S_{Min}, S_{Max}] \,. \tag{2d}$$

We shall discuss later the suitable choices for the lower and the upper bound.

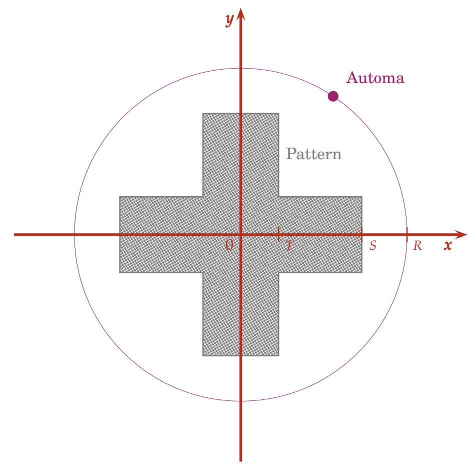


Figure 2: The automa (a point on the external circle) and the object pattern

Clearly, we have that $0 \le T \le S \le R$. Thus, the ratio $\rho = \frac{S}{T}$ has the following restrictions

$$\rho \in [0, 1] \,. \tag{3}$$

The choice of that parameter turn out to be very important, since it will uniquely identify the patterns (together with the size). A discrete set of such ρ's will be selected for this purpose, as it is explained in the following SubSection.

1.1. **Choice of the Patterns.** This SubSection will be devoted to the choice of the total number of the patterns N_o to be categorized. That is a crucial task, since it implies both the choice of some parameters, and the categorization capability of the automa, which we want to maximize.

If we let $\alpha(n) = n\frac{2\pi}{\#Steps} + \alpha_0$ be the angle at the n-th step, because of the special symmetry of the cross, we can restrict our attention to $\alpha(n)$ in the angle-interval $\left[0, \frac{\pi}{4}\right]$, without loss of generalization. We shall denote by $(R\cos\alpha(n), R\sin\alpha(n))$ the position of the automa at the n-th step.

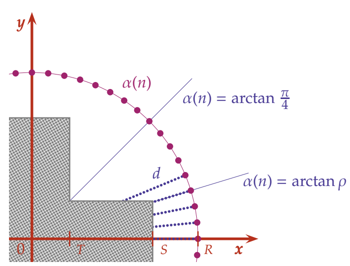


Figure 3: Computation of the distance d.

If we let $d\,(\alpha(n))$ denote the distance of the cross from the circle at the n-th step. It detects the distance from the object by the automa sensor, which we set along the radial direction. We can obtain an explicit definition for the square d, depending on the position (before or after) of that angle with respect to the corner of the cross, that is either

$$0 \le \alpha(n) \le \arctan\rho\,,$$

or

$$\arctan\rho \le \alpha(n) \le \arctan\tfrac{\pi}{4} = 1\,.$$

Namely, $d\,(\alpha(n))$ is the distance from the point $(S, S\tan\alpha(n))$ in the first case, and from $(T\cot\alpha(n), T)$ in the second case. Thus, by Pitagora's Theorem we get

$$d^2\,(\alpha(n)) = \begin{cases} (R\cos\alpha(n) - S)^2 + (R\sin\alpha(n) - S\tan\alpha(n))^2 & \text{if } 0 \le \tan\alpha(n) \le \rho \\ (R\cos\alpha(n) - T\cot\alpha(n))^2 + (R\sin\alpha(n) - T)^2 & \text{if } \rho \le \tan\alpha(n) \le 1 \end{cases},$$

which has the same behavior of $d\,(\alpha(n))$, since $d\,(\alpha(n))$ is positive. For a deeper investigation we need the first derivative of d^2 with respect to the angle $\alpha(n)$,

$$\frac{\partial d^2\,(\alpha(n))}{\partial\alpha(n)} = \begin{cases} -\frac{2S(R\cos\alpha(n)-S)\sin\alpha(n)}{\cos^2\alpha(n)} & \text{if } 0 \le \tan\alpha(n) \le \rho \\ \frac{2T(R\sin\alpha(n)-T)\cos\alpha(n)}{\sin^2\alpha(n)} & \text{if } \rho \le \tan\alpha(n) \le 1 \end{cases}.$$

Thus, recalling the restrictions for the angle-interval, the derivative has the following sign

$$\frac{\partial d^2(\alpha(n))}{\partial \alpha(n)} \geq 0 \iff \begin{cases} R\cos\alpha(n) - S \leq 0 & \text{if } 0 \leq \tan\alpha(n) \leq \rho \\ R\sin\alpha(n) - T \geq 0 & \text{if } \rho \leq \tan\alpha(n) \leq 1 \end{cases} .$$

Therefore, we get

$$\begin{cases} d^2(\alpha(n)) & \text{decreases if } 0 \leq \alpha(n) \leq \arctan\rho \\ d^2(\alpha(n)) & \text{increases if } \arctan\rho \leq \alpha(n) \leq \frac{\pi}{4} \end{cases} ,$$

that means that we shall always find a minimum for $\alpha(n) = \arctan\rho$, whereas the maximum is achieved either for $\alpha(n) = 0$ or $\alpha(n) = \frac{\pi}{4}$. Hence, the distance $d(\alpha(n))$ has the following critical values

$$d_{Min}(\alpha(n)) = d(\arctan\rho) = R - \sqrt{T^2 + S^2} , \tag{4a}$$

$$d_{Max}(\alpha(n)) = \max\left(d(0), d\left(\tfrac{\pi}{4}\right)\right) = \max\left(R - S, R - \sqrt{2}T\right) . \tag{4b}$$

Thus, four cases may occur in that setting, according to the different mutual positions of the two maxima and the minimum. Since they depend explicitly on the value ρ, we shall chose four values by symmetry as follows.

Square Case: The maximum is achieved for 0, and the minumum. The only possible choice is $\rho_{Square} = 1$.

Thick Cross Case: The maxima are achieved for 0 (absolute maximum) and $\frac{\pi}{4}$ (relative maximum), and the minumum for ρ. The admissible range for ρ is $\left(\frac{\sqrt{2}}{2}, 1\right)$, with mean value $\rho_{Thick} = 0.853553$.

Critical Cross Case: The maxima are achieved for 0 and $\frac{\pi}{4}$ (both absolute maximum), and the minumum for ρ. The exact value is $\rho_{Critical} = \frac{\sqrt{2}}{2} = 0.707107$.

Thin Cross Case: The maxima are achieved for 0 (relative maximum) and $\frac{\pi}{4}$ (absolute maximum), and the minumum for ρ. The admissible range for ρ is $\left[0, \frac{\sqrt{2}}{2}\right)$, and by symmetry we shall choose $\rho_{Thin} = 0.56066$.

1.2. **Input Normalization.** In order to let the scale be uniform, the output scale is set to be equal to the input one. Thus, the first input, devoted to the distance perception (denoted by I_0 in Figure 1) is given by the following expression

$$I_0(n) = \frac{\text{Distance Measured by the Automa Sensor}}{R_{Max}} \in [0, 1] , \qquad \forall n = 0, \ldots, \#Steps . \tag{5}$$

The plot of the detected distance at each step describes a wave (see the Figure 4). The first output, devoted to the distance prediction (denoted by I_0 in Figure 1), is in the interval $[0, 1]$ as well, because of the logistic function.

In order to prevent the network from categorizing patterns only by some ``special˝distances (either over or lower come critical bounds), we can impose that every set of wave pattern has a points that fall within a fixed range, let's say

$$d(n) \in [d_{Min}, d_{Max}] , \qquad \forall n = 0, \ldots, \#Steps . \tag{6}$$

Indeed, we may set the range for the parameter S, given by (2d) in such a way that the correspondent range for the distance, given by (6) is *invariant for any of the four cases*. Therefore, none of the figure will be alloed to exceed that range. Namely, depending on the pattern case, the relations are as follows.

Square and Thick Cross: In the square and the thick cross case, we have $R - S \leq d_{Max}$ and $R - \sqrt{1 + \rho^2}\, S \geq d_{Min}$, that is

$$S \in \left[R - d_{Max}, \frac{R - d_{Min}}{\sqrt{1+\rho^2}}\right] . \tag{7a}$$

Critical and Thin Cross: In the critical and the thin cross case, we have $R - \rho\sqrt{2}\, S \leq d_{Max}$ and $R - \sqrt{1 + \rho^2}\, S \geq d_{Min}$, that is

$$S \in \left[\frac{R - d_{Max}}{\rho\sqrt{2}}, \frac{R - d_{Min}}{\sqrt{1+\rho^2}}\right] . \tag{7b}$$

Recalling that the mean value for R is 2.5, we shall consider

$$d_{Min} = 0.5\,, \qquad\qquad d_{Max} = 2\,. \tag{8}$$

Therefore, recalling the choices for the ratios ρ's made above, the corresponding ranges and mean values for the size parameters will be

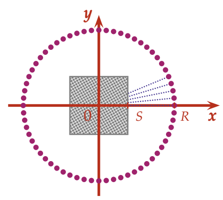
Square: $S \in \left[0.5, \frac{2}{\sqrt{2}}\right]$, with mean value $S_{Square} = 0.957107$;

Thick Cross: $S \in \left[0.5, \frac{2}{\sqrt{1+\rho^2_{Thick}}}\right]$, with mean value $S_{Thick} = 1.010604$;

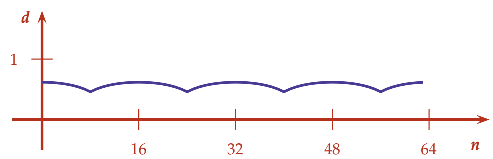
Critical Cross: $S \in \left[\frac{0.5}{\rho\sqrt{2}}, \frac{2}{\sqrt{1+\rho^2_{Critical}}}\right]$, with mean value $S_{Critical} = 1.132993$;

Thin Cross: $S \in \left[\frac{0.5}{\rho\sqrt{2}}, \frac{2}{\sqrt{1+\rho^2_{Thin}}}\right]$, with mean value $S_{Thin} = 1.187561$.

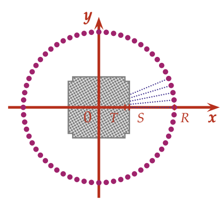
The Figures below represent the four different cases and the relativa wave-plots. In the latter plots the x-axis represent the step n variable, and the y-axis represent the (normalized) distance d at each single step (we chose $\#Steps = 64$ for simplicity).



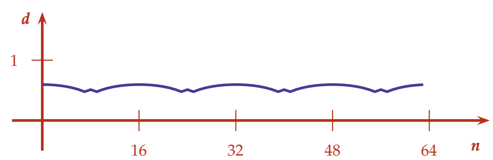
(a) Square Case



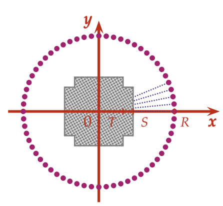
(b) Wave of the Square with $\rho = 1$ and $S_{Square} = 0.957107$



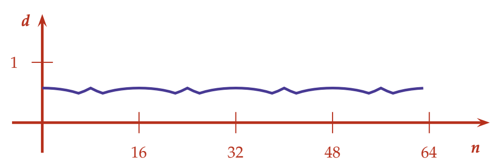
(c) Large Cross Case



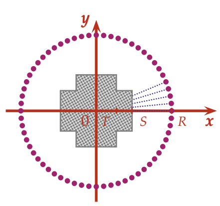
(d) Wave of the Large Cross with $\rho = 0.853553$ and $S_{Thick} = 1.010604$



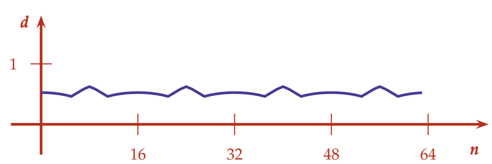
(e) Critical Cross Case



(f) Wave of the Critical Cross with $\rho = 0.707107$ and $S_{Thick} = 1.132993$



(g) Thin Cross Case



(h) Wave of the Thin Cross with $\rho = 0.56066$ and $S_{Thick} = 1.187561$

Figure 4: The four different values of ρ (radius $R = 2.5$)

1.3. **Noises.** There are several types of noises one can introduce in the setting. We choose to restrict our scope to the following three ones.

Angle Error: Error in the angle step $\alpha(n)$. The angle may be longer or shorter than the correct one.
A percentage of 0.05% is set for this type of error. Furthermore, we assume that this type of error is memoryless and thus, there is no cumulation of the automa path errors.

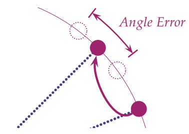


Figure 5: Angle Error

Sensor Error: Error in the sensor. The distance d of the automa from the pattern may be longer or shorter than the correct one. Therefore, the corresponding plot for the wave is not smooth.
A percentage of 0.05% is set for this type of error. Again, we assume that this type of error is memoryless and thus, there is no cumulation.

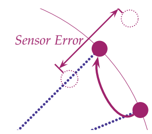


Figure 6: Sensor Error

Ratio Error: Error in the cross ratio. The four discrete values of the ratio ρ may be slightly different from the given one. That is a topological deformation. A percentage of 0.05 is set for this type of error.

Figure 7 shows how the (normalized) wave-plot is effected by these three errors. The dotted line represents the wave-plot in case of absence of noises.

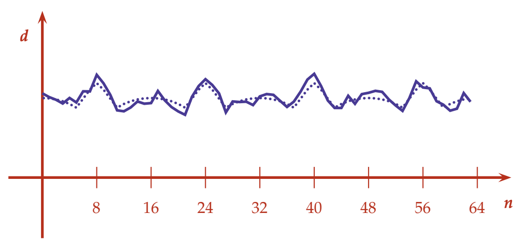


Figure 7: Effects of the Noises in the Thin Cross Wave

It could be interesting to investigate the alternative assumption of a cumulative error, with the necessary assumption of active perception.

1.4. **Goals.** Two kind of goals may be defined for our neural network.

a) Finding special **regularities**. The system should be albe to capture rhythms, lengths and similar phenomena, either with respect to the step scale, and the distance one.
b) Keeping **robustness**. The system should be able to accept a reasonable generalization of the regularities achieved, that is it should be unaffected by noises.

The two goals listed above may be re-phrased in terms of properties of the wave-plot in the xy-plane, where x denotes the step n, and y stands for the distance d. Namely, the first goal is immediately reduced to the periodicity and the positions of the critical points (minima and maxima). The latter task deserves a more detailed analysis, since it is related to suitable transformations of the variables x and y.

For this purpose we consider the following linear transformations of x and y

$$x \longmapsto a_x x + b_x \tag{9a}$$

$$y \longmapsto a_y y + b_y \tag{9b}$$

where a_x, a_y and b_x, b_y are real coefficients. It is interesting to understand the relation of those parameters with the four parameters involved in our setting, given by (2). Suppose that we start with a configuration

$$C = \{\#Steps, \alpha(0), R, S\}\ . \tag{10}$$

Assuming that we fix the ratio ρ, in any case we have the following description.

$\boldsymbol{a_x}$**:** It detects "stretching" along the x axis, and corresponds to inverse variations of the total number of steps, $\#Steps$, or equivalently the speed of the automa. We obtain a new configuration

$$C' = \left\{\tfrac{1}{a_x}\#Steps, \alpha(0), R, S\right\}\ . \tag{11a}$$

It must be positive, and its effects change according to $a_x < 1$ (enlarging), or $a_x > 1$ (shrinking). Figure 8 shows how the wave of the thin cross (see Figure 4(h)) is effected by the transformation $x \longmapsto \frac{1}{2}x$. The dotted line represents the wave before the transformation.

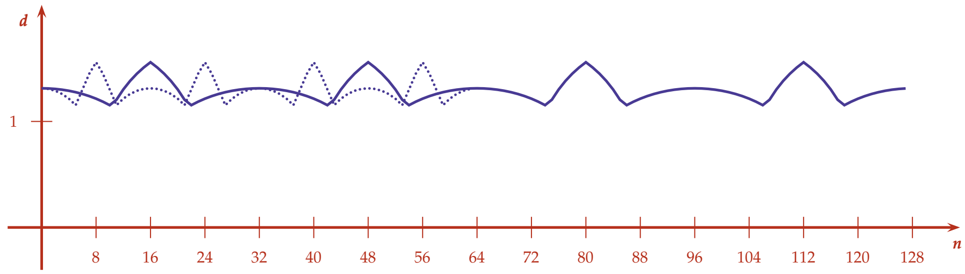


Figure 8: Linear transformation of the Thin Cross Wave with $a_x = \frac{1}{2}$

b_x: It detects the "translations" along the x axis, and corresponds to opposite shifts of the initial angle step $\alpha(0)$ of the automa. We obtain a new configuration

$$C' = \{\#Steps, \alpha(0) - b_x, R, S\} \ . \tag{11b}$$

Figure 9 shows how the wave of the thin cross (see Figure 4(h)) is effected by the transformation $x \longmapsto x + 3$. The dotted line represents the wave before the transformation.

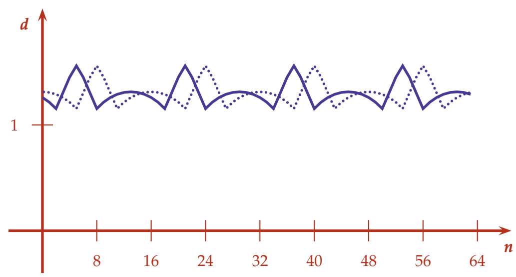


Figure 9: Linear transformation of the Thin Cross Wave with $b_x = 3$

a_y: It detects "stretching" along the y axis, and corresponds to variations of both the size of the cross pattern S, and the radius R of the automa walk. We obtain a new configuration

$$C' = \left\{\#Steps, \alpha(0), a_y R, a_y S\right\} \ . \tag{11c}$$

Figure 10 shows how the wave of the thin cross (see Figure 4(h)) is effected by the transformation $y \longmapsto \frac{1}{3}y$. The dotted line represents the wave before the transformation.

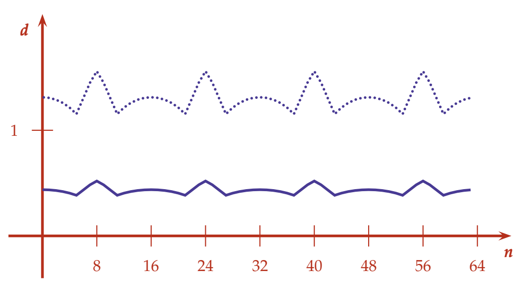


Figure 10: Linear transformation of the Thin Cross Wave with $a_y = \frac{1}{3}$

b_y: It detects the "translations" along the y axis, and corresponds to equal shifts of the radius R of the automa walk. We obtain a new configuration

$$C' = \left\{\#Steps, \alpha(0), R + b_y, S\right\} \ . \tag{11d}$$

Figure 9 shows how the wave of the thin cross (see Figure 4(h)) is effected by the transformation $y \longmapsto y - 0.5$. The dotted line represents the wave before the transformation.

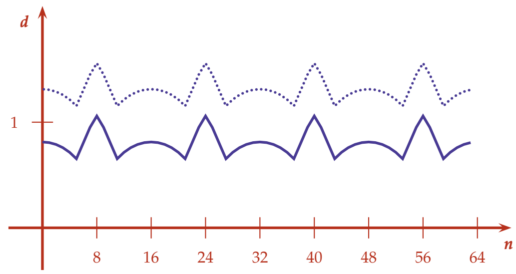


Figure 11: Linear transformation of the Thin Cross Wave with $b_x = -0.5$

References

[1] Cannata, G., and Grosso, E. Active Eye-Head Control. In *Proceedings of the International Conference on Robotics and Automation. Volume 4* (Los Alamitos, CA, USA, May 1994), E. Straub and R. S. Sipple, Eds., IEEE Computer Society Press, pp. 2837–2843. 1

[2] Cecconi, F., and Campenní, M. Recurrent and Concurrent Neural Networks for Objects Recognition. In *Proceedings of the Artificial Intelligence and Applications* (Innsbruck, Austria, 2006). 1

[3] Cecconi, F., and Parisi, D. Neural Networks that Learn to Predict Where Food Is and Also to Eat It. In *Proceedings of the International Joint Conference on Neural Networks* (Washington, USA, 1989). 1

[4] Chella, A., Frixione, M., and Gaglio, S. A Cognitive Architecture for Artificial Vision. *Artificial Intelligence 89*, 1-2 (Jan 1997), 73–111. 1

[5] Eckhorn, R., Reitboeck, H. J., Arndt, M., and Dicke, P. Feature Linking via Synchronization among Distributed Assemblies: Simulations of Results from Cat Visual Cortex. *Neural Computation 2*, 3 (1990), 293–307. 1

[6] Elman, J. L. Finding Structure in Time. *Cognitive Science 14*, 2 (apr 1990), 179–211. 1

[7] Fayman, J. A., Rivlin, E., and Christensen, H. I. A System for Active Vision Driven Robotics. In *Proceedings of the IEEE International Conference on Robotics and Automation* (Minneapolis, USA, apr 1996), pp. 1986–1992. 1

[8] Jang, D., and Choi, H.-I. Moving Object Tracking using Active Models. pp. 648–652. 1

[9] Liu, C.-B., and Ahuja, N. A Model for Dynamic Shape and its Applications. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2* (Washington, USA, 2004), pp. 129–134. 1

[10] Machowski, L. A., and Marwala, T. Representing and Classifying 2D Shapes of Real-World Objects using Neural Networks. In *SMC (7)* (2004), pp. 6366–6372. 1

[11] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning Representations by Back-Propagating Errors. *Nature 323* (1986), 533–536. 1

[12] Shanahan, M. Perception as Abduction: Turning Sensor Data into Meaningful Representation. *Cognitive Science 29*, 1 (2005), 103–134. 1

[13] Sinha, P. R., and Bajcsy, R. K. Robotic Exploration of Surfaces and its Application to Legged Locomotion. In *Proc. IEEE International Conf. on Robotics and Automation* (Nice, France, 1992), pp. 221–226. 1

[14] Sun, B., and Takeda, F. Research on the Neural Network with rbf for Currency Recognition. In *Proc. IASTED International Conf. on Neural Networks and Computational Intelligence* (Grindelwald, Switzerland, 2004), pp. 201–205. 1

[15] Tani, J., and Nolfi, S. Learning to perceive the world as articulated: an approach for hierarchical learning in sensory-motor systems. *Neural Networks 12*, 7-8 (1999), 1131–1141. 1

[16] Thomas, N. J. T. Are theories of imagery theories of imagination? *Cognitive Science 23*, 2 (1999), 207–245. 1

[17] Ziemkea, T., Jirenhedb, D.-A., and Hesslow, G. Internal simulation of perception: a minimal neuro-robotic model. *Neurocoputing 68* (2005), 85–104. 1

Department of Pure and Applied Mathematics, University of L'Aquila, Via Vetoio, 67010 Coppito (AQ), Italy, Tel. no. +39 0862 433189, Fax no. +39 0862 433180
E-mail address: zappacos@univaq.it
URL: http://mat.univaq.it/~zappacos/

Institute of Cognitive Sciences and Technologies, Consiglio Nazionale delle Ricerche, Via S. Martino della Battaglia 44, 00185 Rome, Italy, Tel. no. +39 06 44595231, Fax no. +390644595243
E-mail address: gianluca.baldassarre@istc.cnr.it
URL: http://gral.istc.cnr.it/baldassarre

Institute of Cognitive Sciences and Technologies, Consiglio Nazionale delle Ricerche, Via S. Martino della Battaglia 44, 00185 Rome, Italy, Tel. no. +39 06 44595233, Fax no. +390644595243
E-mail address: stefano.nolfi@istc.cnr.it
URL: http://gral.istc.cnr.it/nolfi/