ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

Can you follow that guy?

Mariacarla Staffa[1], Massimo De Gregorio[2], Maurizio Giordano[3] and Silvia Rossi[1] *

1- Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione – DIETI
Via Claudio 27 – Naples, Italy

2- Istituto di Cibernetica "Eduardo Caianiello" – CNR
Via Campi Flegrei 34, Pozzuoli, Italy

3- Istituto di Calcolo e Reti ad Alte Prestazioni – CNR
Via Pietro Castellino 111, Naples, Italy

Abstract. The problem of tracking moving objects or human beings is a challenging problem in mobile robotics. Knowledge about the position of moving subjects can be used both to improve the behavior of the robotic system, and to perform tasks of monitoring or following. Different methodologies have been applied in literature, using different sensors and techniques for addressing this problem. In this paper we propose a WiSARD-based system approach for tracking either moving robots or human beings.

1 Introduction

In mobile robotics the need of dynamically detect and track objects, and to adapt the robot speed and movements to the dynamics of the surrounding environment, where other robots or human beings can operate, is a crucial issue. The objects tracking problem consists in reconstructing the trajectory of objects along the sequence of images. It is considered as a basic problem in many computer vision applications and it is inherently difficult, especially when applied to real world conditions, where unstructured forms are considered for tracking, real time responses are required for adapting the robot movements in time, computational capabilities are limited to on-board units and where problems of brightness and non-stationary background can affect the performance of the elaboration system.

To overcome this problem, many approaches have been proposed in the literature adopting different types of sensors (video cameras, laser range finder, radar) and methodologies. In many applications, Kalman filters and hence Gaussian distributions are used to track the individual objects [1]. More recently, particle filters have been introduced to estimate non-Gaussian, non-linear dynamic processes [2]. They have been applied with success to different state estimation problems including visual tracking [3].

Here, our aim is to address the problem of making a robot able to follow any object (for example another robot or a human being) that can dynamically change its shape during the tracking. A robot/human tracking method has to be very flexible to identify all the different shape instances but at the same time highly specific, in order not to misclassify the target to follow. In order to have

*Work supported by the European Community, within the FP7 ICT-287513 SAPHARI project and the FP7 ICT-600958 SHERPA project.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

an appropriate system (feature detector) to accomplish this task, we have to deal with noise tolerance (because of dynamic backgrounds and luminance changes), shape, color and scale invariance.

In this paper, we propose a WiSARD–based system [4] as feature detector for tracking heterogeneous moving robots or human beings. This particular weightless neural system has the property of being noise tolerance and is capable of learning step–by–step the new appearance of the moving object on a dynamic background. A similar system has already been developed in [5] with very good results. In fact, WiSARD can be adopted to deploy virtual sensors that, with a limited use of computational resources, can be used on-board for object tracking and dynamical selection of the desired targets to track.

2 WiSARD for Object Tracking

WiSARD systems are based on networks of Random Access Memory (RAM) nodes (neurons). Created by Wilkes, Stonham, and Aleksander in 1984, the WiSARD perceptron was the first artificial neural network machine (and the most representative weightless neural network model – WNN) to be patented and produced commercially [6]. Many other WNN paradigms were proposed and have been surveyed in [4]. A WiSARD takes a set of bits as input, which is then parsed into a set of uncorrelated n-tuples. Each n-tuple is used as a specific address of a RAM–based neuron, in such a way that the input field is completely covered. A WiSARD discriminator, composed by a set of RAM–based neurons, is trained with representative data of a specific class/category. A discriminator recognizes a test pattern via summation of all its RAM neuron outputs. Therefore, the WiSARD model is a multi–discriminator, unidirectional architecture, in summary, a perceptron (see Figure 1).

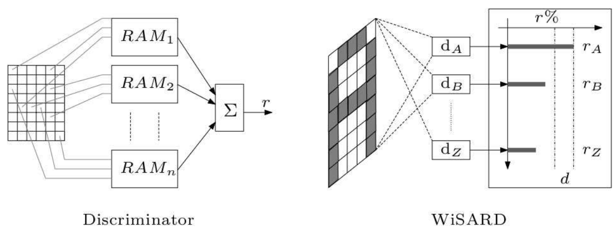


Fig. 1: RAM-discriminator and WiSARD.

The WiSARD system we propose, is formed by a certain number of discriminators each one looking at different parts of the image (frame). Except for the retina (input field) of the central discriminator, all the other retinas are placed all around the initial position. So doing, each discriminator is identified by its relative coordinates. The displacement of all the retinas forms what it is called prediction window.

Let (x, y) be the retina coordinates of the central discriminator ($d_{x,y}$) and h

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

and w the size of the prediction window. The whole set of discriminators are labeled as $d_{x\pm n,y\pm m}$, with $n \in [1, w/2]$ and $m \in [1, h/2]$. The generic discriminator $d_{i,j}$ is going to be responsible for detecting the object in case its new position is identified by (i, j) in the prediction window. The discriminators accept as input binary patterns and the best way to codify the frames (color images) is that of transforming them through the Gray code [7]. In this way we are not loosing the object color information that allows the system to better follow the object.

At the beginning, the system is fed with an image representing the object to be followed (at the moment, the image part representing the object is selected by the user). This image is used to train all discriminators. When the object starts moving, the WiSARD system tries to localize the object through the discriminator responses. The higher is the response the more probable the object is in that part of the prediction window. Once the system localizes the object in the new $(\underline{i}, \underline{j})$ position (that is, discriminator $d_{\underline{i},\underline{j}}$ has the highest response), all discriminators are reset and trained with the image on the $d_{\underline{i},\underline{j}}$ discriminator retina.

3 Case Studies: Robots and People Following

We consider a simple flocking problem, where an agent in the Leader role is moving and a robot (Follower) must keep following the Leader. We adopt a visual servoing approach based on the proposed WiSARD feature detector. We tested the system by considering two different agents to be followed: 1) a wheeled robot and 2) a human being.

The robots used in the case studies are *Pioneer 3-DX*. Each Pioneer is endowed with 16 sonar sensors. The Pioneers wheels actuators accept a linear velocity reference (v) and an angular velocity reference (ω). The control system is implemented as a behavior–based architecture [8] on regular personal computers, via the *Robot Operating System* (ROS) tool for developing robotic systems, and the communication between the computers and the Pioneers is implemented via the *RosAria* wrapper of ROS for mobile robots.

3.1 Robot Behavior Description

In order to accomplish the assigned task the following primitive behaviors have been implemented: *Wander*, *AvoidObstacles* and *FollowMate*.

Wander – This behavior is for generating small random rotations in the robot motion just in the case it loses the Leader.

AvoidObstacles – This behavior is for avoiding collisions with obstacles during the path. In particular, the index h of the sonar s_h detecting the current lower distance from an obstacle is selected and used for regulating the velocity output. Different reactions will be triggered according to the position of the h-th sonar. The smaller the distance from the obstacle, the greater will be this contribution.

FollowMate – This behavior is intended to make a robot follow a particular target (another robot or a human being). The *FollowMate* attitude is accomplished via the proposed WiSARD–based visual servoing approach. The target

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

to be followed is determined at the beginning of the interaction by selecting a region of interest containing the object/person to follow.

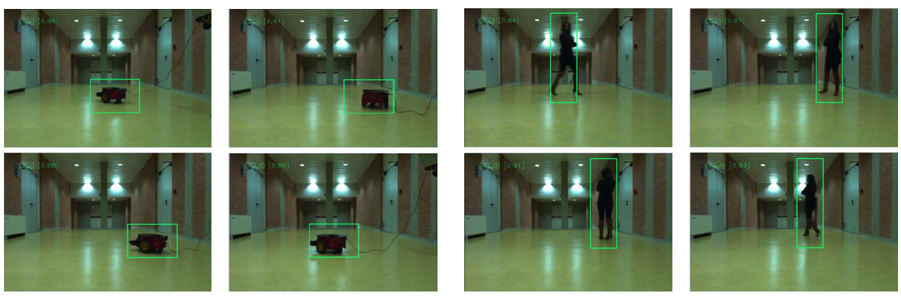
The visual data elaboration follows two steps: i) pre–elaboration of the frames extracted by the video stream; ii) WiSARD frame classification with $w \times h$ discriminators. In particular, we adopted $w = 20$ and $h = 10$ to better detect respectively horizontal and vertical movements.

When WiSARD detects a new position of the object it is supposed to follow, the robot adjusts its velocities (v and ω) according to the discriminator responses. If the best response is given by the discriminator $d_{x,y+j}$, with $j \in m$, this is perceived by the Follower as the robot Leader is moving away. This corresponds to a small increase in linear velocity ($v = v + j \times \epsilon$). On the contrary, a best response from $d_{x,y-j}$ is perceived as the Leader approaching and corresponds to a deceleration of the speed ($v = v - j \times \epsilon$). The $d_{x\pm i,y}$, with $i \in n$, responses solicit respectively right or left movements and produce a small angular velocity in the direction of the perceived movement ($\omega = \omega \pm i \times \epsilon$).

4 Experimental Results

In this section, we report the collected results obtained with a first quantitative assessment of the WiSARD approach to the tracking problem (see Table 1). We first evaluate the WiSARD–based approach as tracking sensor in a case of a Fixed Camera (FC), i.e. without the Follower robot (see Figure 2). We collected results about the quality of the classification, computed as average value of the best classifier responses, and the number of failures by counting the number of times the object exits the prediction window. 10 different 3-minute videos are evaluated and the mean and standard deviation values are reported in Table 1. For the FC, the WiSARD approach is very powerful in terms of right classifications and adaptability. This is mainly due to the intrinsic characteristic of the WNNs of being trained on–line.

Then, we tested the WiSARD–based tracking sensor in experiments of the



(a) Robot Tracking (b) Human Tracking

Fig. 2: Snapshots of the WiSARD–based tracking sensor on a fixed camera.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

duration of 3 minutes where a flocking problem is considered, with 10 undergraduate voluntaries (Robot–Human (RH) Tracking). As for the robotic setting, we tested the approach by repeating the experiment 10 times with the same robot Leader (Robot–Robot (RR) Tracking). Data are reported in Table 1. One can notice that, even if the average value of the WiSARD discriminators responses is higher in the FC Tracking case, it remains reliable even in the other two cases, in which many new challenging aspects have to be taken into account.

In particular, we note that the performance values of case studies 1) and 2) are comparable both in terms of discriminator responses and number of failures. However, we observe a slight performance degradation in terms of discriminator responses in RH Tracking case. This seems to be due to the changes of the human body shape while moving.

	Classifiers	Failures
FC Tracking	0.90 ± 0.02	2.10 ± 0.87
RH Tracking	0.62 ± 0.02	1.10 ± 0.99
RR Tracking	0.78 ± 0.05	1.10 ± 0.73

Table 1: Quantitative evaluation of the WiSARD approach.

In Figure 3, as a matter of example, the Pointing Error trend relative to the execution of the flocking experiment in both case studies is sketched. The Pointing Error is the distance between the prediction window and the image frame center.

The lower the distance, the better the robot tracks and follows the interesting object. One can notice that, being the two plots similar, the proposed system is quite robust in accomplishing flocking tasks where different and heterogeneous subjects have to be followed.

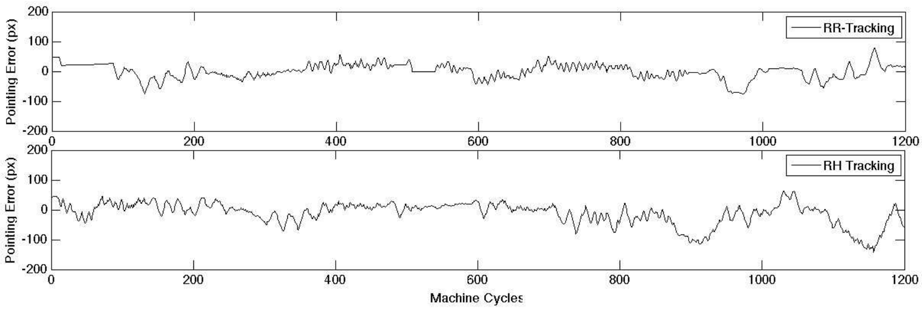


Fig. 3: Pointing Error during experiments of case studies 1) and 2).

5 Conclusions

The proposed WiSARD–based visual servoing approach makes the robot capable of tracking and following different kinds of targets. The on–line training characteristic of the system allows the robot to adapt in real–time to any new situations, such as, a) Leader shape and color changes and b) dynamic switch on a different Leader. This is mainly because the system does not need to have a model of the Leader in advance. On the other side, since our system uses common image filtering techniques, we believe that its performance could benefit from the investigation and development of more accurate filtering techniques. Note that in our system the filtering stage on the original image (frame) is necessary to transform it in compatible input for WiSARD. Furthermore, as future work, in order to improve the performance of the proposed WiSARD–based visual servoing system, we plan to investigate the adoption of a dead reckoning strategy to anticipate the Leader's current position by using its previously determined positions.

ACKNOWLEDGMENT

This work has been partially funded by the European Commission's 7th Framework Programmes as part of the project SAPHARI under grant 287513, and RODYMAN under ERC-grant agreement no. 320992.

References

[1] N. Gordon. A hybrid bootstrap filter for target tracking in clutter. *Aerospace and Electronic Systems, IEEE Transactions on*, 33(1):353–358, 1997.

[2] M. Pitt and N. Shephard. Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association*, 94:446, 1999.

[3] M.J. Black and A.D. Jepson. A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions. In *LNCS 1406*, pages 909–924. 1998.

[4] I. Aleksander, M. De Gregorio, F.M.G. França, P. Lima, and H. Morton. A brief introduction to weightless neural systems. In *ESANN*, pages 299–305, 2009.

[5] H.L. França, J.C.P. da Silva, O. Lengerke, M.S. Dutra, M. De Gregorio, and F.M.G. França. Movement persuit control of an offshore automated platform via a ram-based neural network. In *ICARCV*, pages 2437–2441. IEEE, 2010.

[6] I. Aleksander, W. V. Thomas, and P. A. Bowden. WISARD a radical step forward in image recognition. *Sensor Review*, 4:120–124, 1984.

[7] F. Jianying, L. Linchao, G. Yang, Z. Zeliang, Y. Lei, and L. Wei. Research on color gray code encoding and color components correction in 3d measurement for color object. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(5):217–226, 2013.

[8] R.C. Arkin. *Behavior Based Robotics*. MIT Press, 1998.