ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

A WiSARD-based multi-term memory framework for online tracking of objects

Daniel N. do Nascimento[1], Rafael L. de Carvalho[1,3], Félix Mora-Camino[4], Priscila V. M. Lima[2], Felipe M. G. França[1] *

1 – COPPE, 2 – NCE, Universidade Federal do Rio de Janeiro, BRAZIL

3 - Universidade Federal do Tocantins, UFT, BRAZIL

4 - Ecole Nationale de l'Aviation Civile - Laboratoire d'Automatique, FRANCE

Abstract. In this paper it is proposed a generic object tracker with real-time performance. The proposed tracker is inspired on the hierarchical short-term and medium-term memories for which patterns are stored as discriminators of a WiSARD weightless neural network. This approach is evaluated through benchmark video sequences published by Babenko et al. Experiments show that the WiSARD-based approach outperforms most of the previous results in the literature, with respect to the same dataset.

1 Introduction

The task of tracking objects with real-time requirement has many useful applications. In this context, SanMiguel *et al.* [1] proposed a framework for estimation of the quality in video tracking algorithms, which features the capability of evaluating video trackers with multiple failures and recoveries over long sequences. On the other hand, Percini and Del Bimbo [2] presented an object tracking method where multiple instances of scale invariant local features were considered. Their method used a non parametric learning algorithm based on the transitive matching property, presenting state of the art tracking performance on public available benchmark datasets. Regarding the task of tracking an object, given its position in the first frame of a sequence, authors in [3] considered a Multiple Instance Learning algorithm for which the training data is provided through labeled bags, instead of labeled instances.

Application of weightless neural network models on tracking tasks were reported in [4], where the WiSARD model was successfully used by an artificial vision system in order to follow the cadence of ships. The results hinted the possibility of applying the mechanism to model the movement of an observed vessel. Furthermore, in [5] promising results were shown concerning the use of the WiSARD weightless neural network on a real-time tracker application. In this work, the same benchmark and metrics adopted in [3] were considered as means of evaluating the accuracy of the proposed tracker. In contrast with the adopted methodology in [2], this work makes use of a minimum number of features, with the objective of maximizing tracking speed. In summary, the proposed tracker has the objective of following an object identified in the first frame of a video sequence.

*The authors acknowledge the Brazilian Agencies CAPES and CNPq for partially funding this research.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

This paper is organized as follows. The WiSARD model is reviewed in Section 2, followed by Section 3, which describes the short- and medium-term hierarchic memory tracker. Experimental setup and results are presented in Section 4. Finally, in section 5, some conclusions and future improvements are pointed out.

2 WiSARD

WiSARD is a weightless neural network and its name stands for Wilkie, Stonham and Aleksander's Recognition Device [6]. The WiSARD is structured as a network of discriminators, which are composed of RAM-based neurons. Patterns are represented as the energized paths from the network input to each of the RAM-based neurons of the discriminators. Each RAM has a number of input entries given by the binary address formed by its corresponding input subpattern. In training mode, an addressed pattern is stored in a RAM position as an integer value different from zero; non-addressed entries remain zero. Besides, in classification mode, each discriminator outputs the number of addressed RAM positions, for which the address was energized in training mode. Figure 1 illustrates the architecture of the WiSARD's discriminator.

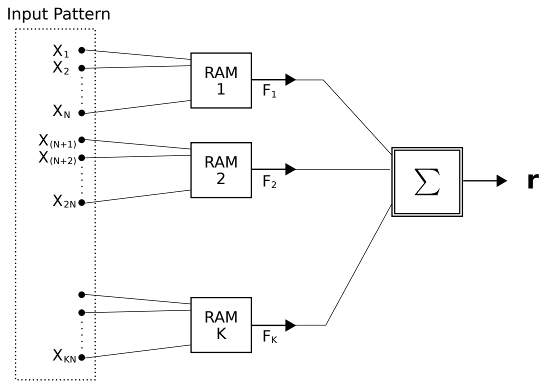


Fig. 1: Discriminator: a basic unity for the WiSARD model [6].

3 Hierarchic memory tracker

Inspired by the human memory hierarchy, the proposed tracker is based on the concept of short- and medium-term memories. It is assumed that the shape changing of a object is seen as a new pattern to be learnt. For each pattern of the followed object, the proposed tracker stores a number of discriminators, each one representing a pattern learned in different moments of the tracking process. Thus, the hypothesis is that it is possible to keep tracking the object even if it changes its shape or becomes occluded for a period of time.

In the beginning of the process, the location of the object in the first frame is used as an input to the tracker, which trains the first discriminator and stores it in the hierarchic memory. For the next frames, the discriminator is used to

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

find the object at the scene, locally searching around the last object's location. The discriminator returns a score to each position inside the searched region, and the position that returns the higher score is assumed to be the location of the object in the current frame. This process goes on until the classification score reaches a *pattern threshold*. When the score falls below this threshold, the tracker assumes that a new discriminator has to be trained in order to learn the new object shape. The tracker then proceeds to storing the current discriminator into the medium-term memory, and training a new discriminator to assume that position into the short-term memory.

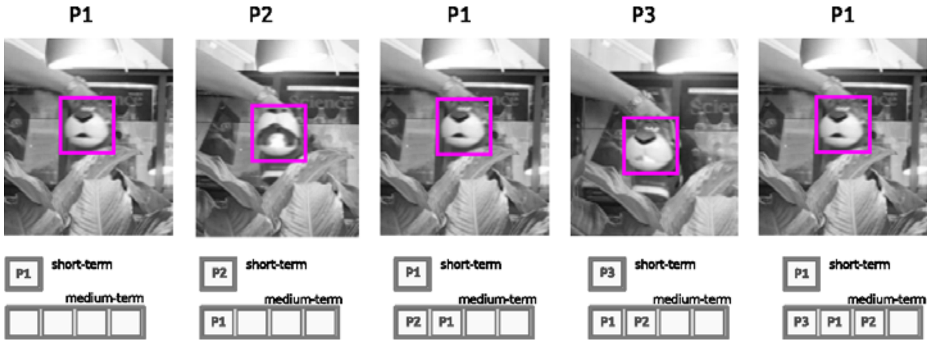


Fig. 2: Hierarchic Memory example: At first, the discriminator P1 is used to find the object; in sequence, a new discriminator P2 is trained and placed in the first position; then, if the discriminator P1 returns the best score, it goes to the first place of the queue. In a future frame, P3 is trained and placed in the first position, then, if discriminator P1 returns the best score, it goes to the first position.

For each new frame, the tracker searches for the object pattern into both memory queues. The discriminator that gives the best score is chosen to represent the object location at the current frame, and that discriminator is transferred to the first position of the queue. Both queues have a maximum number of discriminators they can store. When this maximum number is reached, the discriminator located in the last position is dropped. This process guarantees that the most recently seen patterns are maintained in the hierarchic memory. Using this strategy, the discriminator that has not been used for the longest time, is naturally discarded when it is necessary to release memory to allocate a new discriminator. Figure 2 illustrates an example of allocation at the hierarchic memory with capacity to store four discriminators.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

4 Experimental setup and results

We ran the WHMTracker with default and tuned parameters in the same set of videos[1] examined in [3]. The video clips names and the corresponding default and tuned parameters are shown in Table 1. All videos are in gray scale and present some problematic situations for a tracking system to handle, such as occlusion and shape changing over time. Before training a discriminator, the cropped image of the object, given by the bounding box, is binarized. For this purpose, the mean value of luminance is used as threshold. This process is employed while the tracker is searching for the object around a local neighborhood.

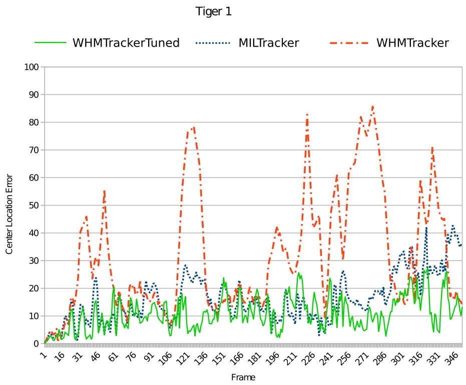
Table 1: Default and tuned parameters used in each tested video clip. Video clips identified with * indicate that a background extraction procedure is also part of the parameters.

Video	*Bits*	*New disc.*	*Memory Size*	*Search area*
Default params.	*5*	*0.7*	*6*	*12*
Tiger1*	default	0.35	20	14
Tiger2*	default	0.35	20	16
Occluded Face	3	0.5	10	10
Occluded Face 2	3	0.5	10	10
David Indoor	6	default	default	10
Sylvester	3	0.8	default	5

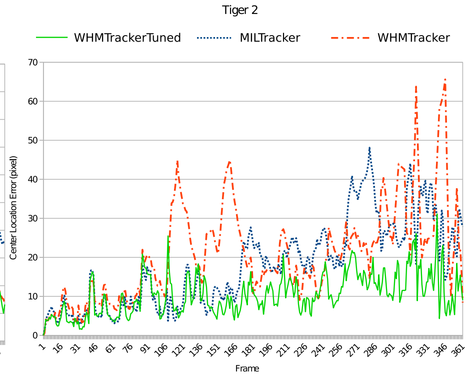
Table 2: Average Center Location Errors (in pixels). Values marked with '*' indicate the best performance and boldfaced ones represent the second best performances.

Video Clip	*MILTrack*	*WHMTrack*	*WHMTrackTuned*	*FPS*
Sylvester	**11**	22	8*	87
David Indoor	23	**11**	8*	22
Occluded Face	**27**	**27**	12*	17
Occluded Face 2	20	**16**	9*	28
Tiger 1	**16**	33	11*	45
Tiger 2	**18**	21	10*	43
Coupon Book	**15**	4*	4*	21

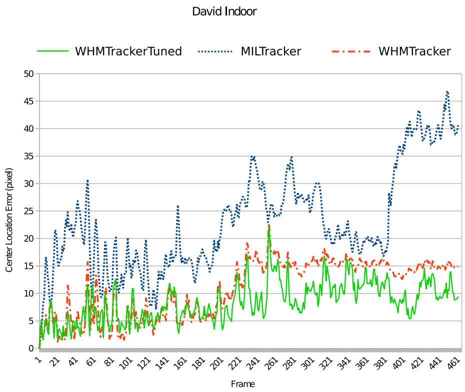
[1]The set of videos is available in: `http://vision.ucsd.edu/~bbabenko/project_miltrack.html`



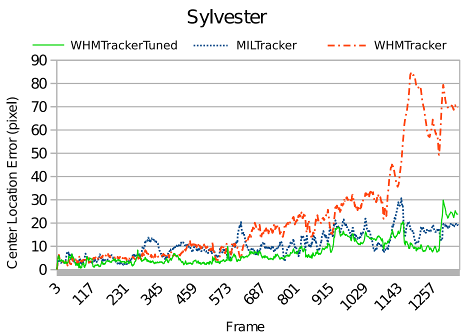
(a)



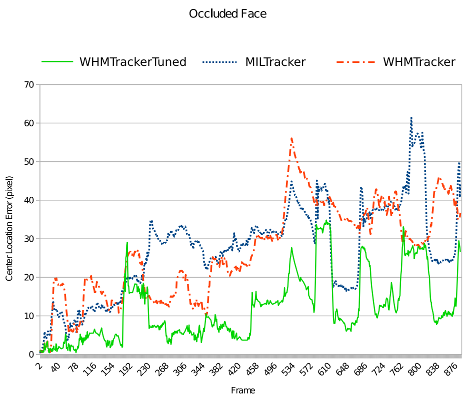
(b)



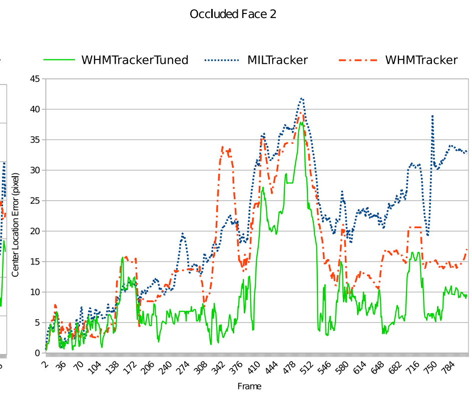
(c)



(d)



(e)



(f)

Fig. 3: Figures 3a , 3b, 3c, 3d, 3e and 3f show a comparison among the tracker in [3], WHMTracker and Tuned WHMTracker for *Tiger1*, *Tiger2*, *DavidIndoor*, *Sylvester*, *OccludedFace* and *OccludedFace2* video clips, respectively.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

In order to compare the results of the proposed tracker with the ones in [3], the average bounding box center error was adopted. Each video clip includes the associated ground truth data, which gives the position and size of the object from 5 to 5 frames. The same linear interpolation as in [3] was used to get the bounding box information for each frame. In addition, the tracker was executed 5 times for each video and the average error of the center location error was computed. Figure 3 shows the results for the set of video clips. Each plot has three pieces of information: the MILTrack result as well as the WHMTracker with and without tuned parameters. Table 2 shows a comparison between the results herein obtained and those in [3][2].

5 Final Remarks

In this paper, an object tracker for real time applications was presented. The tracker uses a hierarchic memory architecture in order to store a queue of object patterns represented by discriminators of the WiSARD model. This memory architecture model was important to overcome problems such as occlusion, because a memory of a past seen object is stored and it is used as soon as the object becomes visible again.

As shown by the experiments, the proposed tracker is able to surpass the results presented in [3], using tuned parameters. The online training of a new discriminator representing a new object pattern was possible due to the WiSARD architecture, which allows for one shot learning. The main shortcoming of the proposed solution is the parameterization search. Future improvement includes search over the environment and object properties in order to propose a solution for auto tuning the tracker parameters.

References

[1] J. SanMiguel, A. Cavallaro, and J. Martínez. Adaptive online performance evaluation of video trackers. *IEEE Transactions on Image Processing*, 21(5):2812 – 2823, 2012.

[2] Federico Pernici and Alberto Del Bimbo. Object tracking by oversampling local features. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, in press, 2014.

[3] Boris Babenko, Nakul Varma, Piotr Dollár, and Serge Belongie. Multiple instance learning with manifold bags. In *International Conference on Machine Learning (ICML)*, Bellevue, WA, 2011.

[4] H.L. França, J.C.P. da Silva, M. De Gregorio, O. Lengerke, M.S. Dutra, and F.M.G. França. Movement persuit control of an offshore automated platform via a ram-based neural network. In *Control Automation Robotics Vision (ICARCV), 2010 11th International Conference on*, pages 2437 –2441, dec. 2010.

[5] R.L. Carvalho, D. S. Carvalho, F. Mora-Camino, P. V. Lima, and F. M. G. França. Online tracking of multiple objects using wisard. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 541–546, April 2014.

[6] Igor Aleksander and Helen Morton. *An introduction to Neural Computing.* Thomson Computer Press, Berkshire House, London, UK, second edition edition, 1995.

[2]Some tracker demos are available in the companion website which is located at `http://labia.cos.ufrj.br/publicacoes/artigos/weightless-hierarchy-memory-tracker`