ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

WiSARD^{rp} for Change Detection in Video Sequences

Massimo De Gregorio[1] and Maurizio Giordano[2]

1 - Istituto di Scienze Applicate e Sistemi Intelligenti “E. Caianiello” (ISASI – CNR)

2 - Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR – CNR)

Abstract. Weightless neural networks (WNNs) have been successfully used as learners and detectors of background regions in video processing, as they feature fast learning algorithm, noise tolerance and an incremental update of learnt knowledge, also referred to as *online training*. These features make WNNs suitable and effective to be used for change detection in scenarios in which environmental changes (light, camera view, cluttered background) and moving objects force the modeling of background regions to change continuously, and in drastic ways. In this paper, we present a change detection method in video processing that uses a WNN, called WiSARD^{rp}, as underlying learning mechanism, equipped with a reinforcing/weakening scheme, that builds and continuously updates a model of background at pixel-level. The performance of the proposed change detection method is evaluated on the ChangeDetection.net video archive.

1 Introduction

Change (motion) detection is one of the most challenging preprocessing tasks in several video processing applications. The majority of change detection methods relies on calculating the evolution of a background model during the video, and change/motion detection is accomplished by comparing (or subtracting) the updated background model to (from) the current frame. Several approaches were investigated in this field: from statistical to clustering techniques, from neural networks to wavelet transformations, from machine learning to bio-inspired approaches. Some of the most recent surveys of existing background modeling and subtraction methods are [1, 2].

In this work, a change detection method based on weightless neural networks (WNNs) is proposed. The method models the background at pixel-level by learning pixel color changes over time. In our approach, the pixel model is represented by the state of a WNN called WiSARD^{rp} [3]. The background model learned by neurons is used to classify pixels as either background or foreground. The neural network associated to each pixel, while modeling the background, features highly adaptivity and noise-tolerance, which are inherent characteristics of WiSARD^{rp}.

While, in weighted neural networks, training on samples consists in a time-spending forward-backward recursive optimization process of neurons weights, in WNNs training at each new sample is just an incremental and one-way update process of neuron memories across network layers. This aspect makes WNNs very fast in learning. In addition, the incremental learning allows the implementation of different and flexible forgetting mechanisms of knowledge about old (out-of-date) data during training.

The main contribution of this work is the investigation of flexible and adaptive learning schemes in neural networks used to model pixel background along the video duration. Indeed, due to its incremental learning mode, in WiSARD^{rp} learned knowledge

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

can be reinforced as well as weakened step by step. This feature makes these systems capable of rapidly adapting the model about a data domain to fast/slow changes in time in that domain. An example of such a domain is foreground detection in videos naturally and/or artificially subject to changes in luminance, movement, shadowing, *etc*.

With the exception of our previous works [4, 5], no method for background modeling based on WNNs has ever been proposed in the past. Self-organizing neural networks [6], general regression neural networks [7], self-organizing maps [8], and adaptive resonance theory neural networks [9] are examples of weighted neural network approaches to background modeling.

To prove our ideas we developed *CwisarDRP*, a change detection software exploting the learning mechanism of WiSARD^{rp} as modeling technique for video background. *CwisarDRP* demonstrated to provide good performance in the average on the ChangeDetection.net 2014 (CDnet 2014) video dataset [10]. Our results and their ranking with respect to state-of-the-art methods are published on the CDnet 2014 web site.

2 Related works for the benchmarking

In what follows, some change detection methods that ranked top scores in the CDnet 2014 competition are discussed. These systems represent an essential benchmark for carrying out an objective comparative analysis of the performance of our method.

FTSG [11] fuses in one system two independent foreground detection methods: 1) a motion detection method based on spatio-temporal tensor formulation and 2) a background subtraction method based on split Gaussian models. FTSG uses a rule-based system to reduce errors due to noise and illumination changes, while an object-level recognition module supervises foreground detection by removing stopped objects.

SuBSENSE [12] method relies on spatio-temporal features of single pixels, namely LBSP [13], whose processing depends on a threshold modulated by the observable pixel color intensity. This makes the method robust with respect to illumination changes and allows to cope with shadow effects. SuBSENSE is equipped with an adaptability mechanism of segmentation parameters that uses the information gathered in monitoring of the background model's evolution as feedback to the system parameter tuning.

PAWCS [14] method characterizes and monitors background representations at pixel-level by using a word-based approach without clustering. The basic idea is to register the occurrences of pixels over time as "background words" in local dictionaries using color and texture information, that is LBSP [13] features with colors. Background words are considered good representations when they reoccur often, while infrequent background words are discarded and replaced by better alternatives.

IUTIS-3 [15] uses genetic programming to automatically determine the best selection and combination of change detection algorithms in a set of detectors.[1] IUTIS-5 [15] method likely[2] is supposed to use the same genetic algorithm but applied to the current three (or more than three) top performant methods.

In SharedModel [16], authors introduce shareable models of Gaussian Mixture Models. Each pixel dynamically searches for the best matched model in its neigh-

[1] SuBSENSE [12], FTSG [11], and CwisarDH [5], the first three ranked methods at CDnet 2014 challenge.

[2] No additional information on this new version of the method is available at the date of this submission.

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

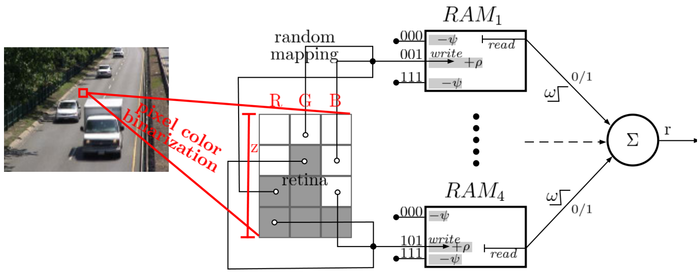

Fig. 1: A WiSARD^{rp} discriminator

borhood, which is then used for classification. The sharable models are built for both background and foreground.

3 The WiSARD^{rp} weightless neural model

This work relies on the WiSARD^{rp} neural model of computation,[3] introduced for the first time in [3] and here formalized and completed in its design (see Figure 1). The core component of this neural network is the *discriminator*, a layer of n-tuple neurons, each one mapping a set of n bits pseudo-randomly extracted from a binary input pattern, also called *retina*. An n-tuple neuron is basically a RAM with 2^n memory cells (and address lines). An n-tuple of bits extracted from the retina is a stimulus for one of these lines, used to write (learn) or read (analyze) neuron contents. This is why weightless neurons are also called n-tuple RAM neurons. Given a binary pattern of size S, the so-called *retina*, it can be learned/classified by a discriminator having m RAMs with 2^n cells each, such that S=$m{\cdot}n$. Discriminators are trained to learn binary patterns belonging to a particular class. The WiSARD^{rp} is the set of so-trained discriminators whose similarity responses on new samples are combined in a winner-take-all scheme.

In the *learning phase*, the RAM cell stimulated by an n-tuple extracted from the retina is incremented by ρ (reward). At the same time, not-stimulated cells are decreased by ψ to zero value (punishment).

In the *classification phase*, each neuron outputs 1 if the value of its stimulated cell is greater than a threshold, namely ω, otherwise it outputs 0. This firing condition makes neurons to contribute to the classification response only when the counter of the sub-pattern under test overcomes a threshold. Let m^* be the number of contributing neurons, then r=$\frac{m^*}{m}$ is the *discriminator response*, which is a measure of the "similarity" between the input pattern and the "current knowledge" of neurons (i.e. the model) about (of) patterns seen during learning.

The *reward & punishment* strategy allows to size the time a sub-pattern is completely unlearned or will remain in the network learned knowledge.

[3]An extension of the WiSARD (Wilkes, Stonham and Aleksander Recognition Device) [17].

	Total rank	Avg Ranking	Avg Recall	Avg Specificity	Avg F-measure	Avg Precision
IUTIS-5	**2.18**	**2.71**	0.7849	0.9948	**0.7717**	**0.8087**
IUTIS-3	5.45	5.00	0.7779	0.9940	0.7551	0.7875
PAWCS	6.36	4.71	0.7718	**0.9949**	0.7403	0.7857
SuBSENSE	7.82	7.57	**0.8124**	0.9904	0.7408	0.7509
SharedModel	8.64	7.00	0.7657	0.9922	0.7283	0.7696
FTSG	8.82	9.14	0.7657	0.9922	0.7283	0.7696
SaliencySubsense	9.55	10.43	0.7714	0.9914	0.7176	0.7628
M4CD 2.0	9.73	12.57	0.7416	0.9923	0.7129	0.7754
Superpixel SBS	9.91	10.71	0.7416	0.9923	0.7129	0.7754
CwisarDRP	10.27	10.86	0.7062	0.9947	0.7095	0.7880

Table 1: First 10 ranked methods on CDnet 2014 competition on November 2016

4 Change detection by WiSARD^{rp}

The change detection software[4] proposed in this paper is called *CwisarDRP*.[5] Its core logic is based on the following assumptions:

Color encoding – The pixel color is represented by three non-negative numbers, namely *color channels* (in RGB, HSV, or Lab space), in the range $[0, 255]$. The three channels are scaled and discretized to the integer range $[0, 1, \ldots, z\text{-}1]$, thus representing a color with a binary pattern of size $3 \times z$ (retina) that can be fed as input to WiSARD^{rp}.

Change detection – The system does not need to be trained in advance: from the beginning of the video, one step of classification is always followed by one step of learning. Therefore, the system is able to classify pixels as belonging to the background or foreground from the very first frames. On the basis of the learned knowledge of pixel color history stored in neurons, the system gives a classification response for the current pixel, that is a measure of its similarity to the color knowledge acquired by the WiSARD^{rp} by means of a continuous training. This response is then compared to the threshold σ to state whether the pixel in the current frame is detected as background.

Background modeling – The pixel background model is represented by the snapshot of RAM contents during time, which can be interpreted as the frequency histogram of all binary sub-patterns occurred during training modulated by the characteristic WiSARD^{rp} forgetting mechanism [18]. WiSARD^{rp} training, i.e. background modeling at pixel-level, is continuous, in that it is applied at each frame and to all pixels regardless of the previous classification response. Nevertheless, a short transient phase (ω frames) is required for the system to produce an initial background model.

5 Experimental study

Experiments have been carried out on the CDnet 2014 dataset by running *CwisarDRP* with the same parameter setting, that is: z=128, ρ=1, ψ=1, ω=30, σ=50, n=16. In the aforementioned configuration, we allocated 24 neurons for each pixel, thus leading, for example, to around 9.5M neurons for a 720×576 video resolution. With this resolution, an OpenMP C++ implementation of *CwisarDRP* performs at 9.7 fps on a 2GHz Intel

[4] Available for download at https://github.com/giordamaug/CwisardRP

[5] *CwisarDRP* is an evolution of previous systems *CwisarD* (ranked 4^{th} in the CDnet 2012 competition on May 2013) [4] and *CwisarDH* (ranked 3^{th} in the CDnet 2014 competition on June 2014) [5]. However, these systems needed a set of preliminary frames to be trained with before starting classifying.

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

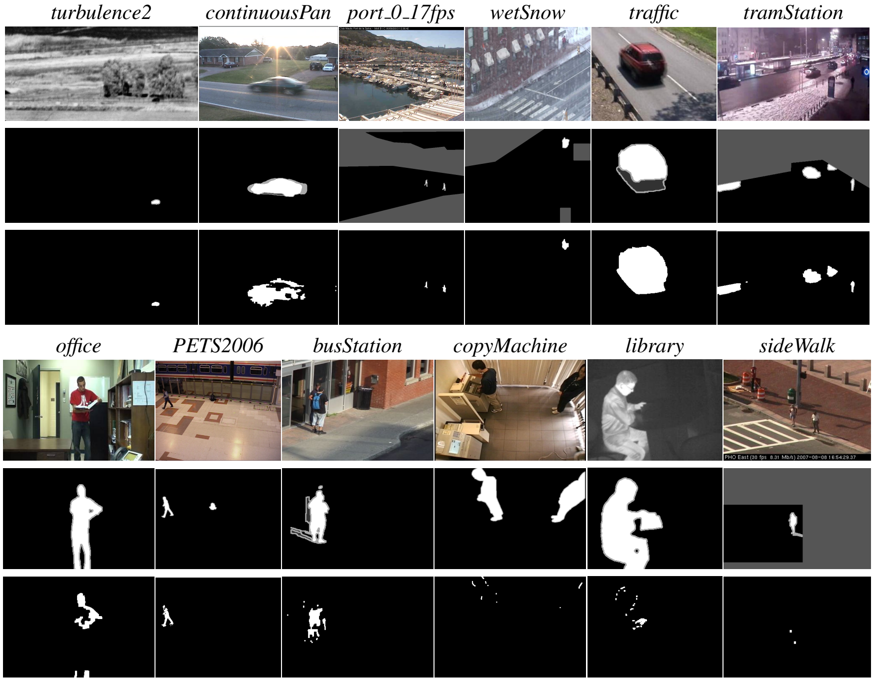

Fig. 2: Best (upper half) and worst (bottom half) results on CDnet 2014 video archive

QuadCore i7 with 16 GB RAM. In Table 1, the average metrics across all video categories and the ten top ranked methods[6] of CDnet 2014 are shown. As it can be noticed, *CwisarDRP* is ranked 2^{nd} in *Average Precision* and 4^{th} in *Average Specificity*.

The results of *CwisarDRP* are very encouraging in spite of its semplicity, and we did not expect to be such a good competitor in the CDnet 2014 challenge. Although our method classified 10^{th} in the overall ranking, it has to be noticed that, if we consider only the *base change detectors*, we are the 7^{th} best method. Indeed, IUTIS-5 and IUTIS-3 are rather *meta detectors*, since they are an "ensemble" of base detectors combined in an optimal and self-adjusting way by a GP algorithm. Because *CwisarDRP* constantly updates the background model frame by frame, it can easily handle videos such those in *PTZ*, *Turbulence*, *Bad Weather*, and *Dynamic background* categories in which background continuously changes. Furthermore, background is perfectly detected by the system in the *continuosPan* and *wetSnow* videos (see Figure 2). *CwisarDRP* does not have the same performance in those videos where objects or persons appear in the scene, stand still for a while and then move away (*i.e.*, *office, copyMachine*). In fact, the system absorbs objects or persons in the background model after ω frames. In the meanwhile, ghosts and false ghosts appear in the elaborated frames.

[6]Although included in Table 1, SaliencySubsense, M4CD 2.0, and Superpixel SBS methods are not discussed in the related works because there are no references on the CDnet 2014 website.

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

6 Conclusions

The change detection method presented in this work, called *CwisarDRP*, is a simple algorithm straight working at pixel-level and exploiting a WNN to model pixel background without any particular pre- and post-processing of image frames. Although its simplicity, it ranked 10^{th} on the CDnet 2014 competition and, at the moment, is the best neural network approach to change detection. This method works very well in those domains where: 1) there is no opportunity to train the system in advance; 2) the scene has to be analyzed 24 hours a day; 3) the user can define when either an object or a person can be considered belonging to the new background (ω).

References

[1] Yong Xu, Jixiang Dong, Bob Zhang, and Daoyun Xu. Background modeling methods in video analysis: A review and comparative evaluation. *CAAI Trans. on Intelligence Technology*, 1(1):43 – 60, 2016.

[2] Thierry Bouwmans. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11-12:31–66, 2014.

[3] Massimo De Gregorio, Maurizio Giordano, Silvia Rossi, and Mariacarla Staffa. Experimenting WNN support in object tracking systems. *Neurocomputing*, 183:79 – 89, 2016. Weightless Neural Systems.

[4] M. De Gregorio and M. Giordano. A WiSARD–based approach to CDnet. In *BRICS-CCI CBIC 2013*, pages 172–177, 2013.

[5] M. De Gregorio and M. Giordano. Change detection with weightless neural networks. In *Proc. of 2014 IEEE CVPRW*, pages 409–413, 2014.

[6] L. Maddalena and A. Petrosino. A self–organizing approach to background subtraction for visual surveillance applications. *Image Processing, IEEE Trans. on*, 17(7):1168–1177, 2008.

[7] Dubravko Culibrk et al. A neural network approach to bayesian background modeling for video object segmentation. In *Proc. of VISAPP 2006*, pages 474–479, 2006.

[8] Z. Zhao, X. Zhang, and Y. Fang. Stacked multilayer self–organizing map for background modeling. *Image Processing, IEEE Transactions on*, 24(9):2841–2850, 2015.

[9] Rafael Marcos Luque, Enrique Domínguez, Esteban J. Palomo, and José Muñoz. An art–type network approach for video object detection. In *ESANN 2010*, pages 423–428, 2010.

[10] Yi Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar. Cdnet 2014: An expanded change detection benchmark dataset. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, pages 393–400, June 2014.

[11] Rui Wang, F. Bunyak, G. Seetharaman, and K. Palaniappan. Static and moving object detection using flux tensor with split gaussian models. In *IEEE CVPRW2014*, pages 420–424, June 2014.

[12] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin. Subsense: A universal change detection method with local adaptive sensitivity. *Image Processing, IEEE Transactions on*, 24(1):359–373, Jan 2015.

[13] G.-A. Bilodeau, J.-P. Jodoin, and N. Saunier. Change detection in feature space using local binary similarity patterns. In *Int. Conf. on Computer and Robot Vision (CRV)*, pages 106–112, May 2013.

[14] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin. A self-adjusting approach to change detection based on background word consensus. In *IEEE WACV 2015*, Jan 2015.

[15] Raimondo Schettini Simone Bianco, Gianluigi Ciocca. How far can you get by combining change detection algorithms? *IEEE Transactions on Image Processing*, May 2015. Submitted.

[16] Yingying Chen, Jinqiao Wang, and Hanqing Lu. Learning sharable models for robust background subtraction. In *2015 IEEE International Conference on Multimedia and Expo*, pages 1–6, June 2015.

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

[18] Massimo De Gregorio and Maurizio Giordano. Background modeling by weightless neural networks. In *New Trends in Image Analysis and Processing – ICIAP 2015*, LNCS, 9281, pages 493–501. 2015.