Face Recognition with VG-RAM Weightless Neural Networks

Alberto F. De Souza[1], Claudine Badue[1], Felipe Pedroni[1], Elias Oliveira[2], Stiven Schwanz Dias[1], Hallysson Oliveira[1], and Soterio Ferreira de Souza

[1] Departamento de Informática
[2] Departamento de Ciência da Informação
Universidade Federal do Espírito Santo
Av. Fernando Ferrari, 514, 29075-910 - Vitória-ES, Brazil
alberto@lcad.inf.ufes.br

Abstract. Virtual Generalizing Random Access Memory Weightless Neural Networks (VG-RAM WNN) are effective machine learning tools that offer simple implementation and fast training and test. We examined the performance of VG-RAM WNN on face recognition using a well known face database—the AR Face Database. We evaluated two VG-RAM WNN architectures configured with different numbers of neurons and synapses per neuron. Our experimental results show that, even when training with a single picture per person, VG-RAM WNN are robust to various facial expressions, occlusions and illumination conditions, showing better performance than many well known face recognition techniques.

1 Introduction

Computerized human face recognition has many practical applications, such as access control, security monitoring, and surveillance systems, and has been one of the most challenging and active research areas in computer vision for many decades [1]. Even though current machine recognition systems have reached a certain level of maturity, the recognition of faces with different facial expressions, occlusions, and changes in illumination and/or pose is still a hard problem.

A general statement of the problem of machine recognition of faces can be formulated as follows: given an image of a scene, identify or verify one or more persons in the scene using a database of faces. In identification problems, given a face as input, the system reports back the identity of an individual based on a database of known individuals; whereas in verification problems, the system confirms or rejects the claimed identity of the input face. In both cases, the solution typically involves segmentation of faces from scenes (face detection), feature extraction from the face regions, recognition, or verification. In this work, we examined the recognition part of the identification problem only.

Many methods have been used to tackle the problem of face recognition. One can roughly divide these into (i) *holistic* methods, (ii) *feature-based* methods, and

(iii) *hybrid* methods [1]. Holistic methods use the whole face region as the raw input to a recognition system. In feature-based methods, local features, such as the eyes, nose, and mouth, are first extracted and their locations and local statistics (geometric and/or appearance) are fed into a classifier. Hybrid methods use both local features and the whole face region to recognize a face.

Among holistic approaches, eigenfaces [2] and fisher-faces [3, 4] have proved to be effective in experiments with large databases. Feature-based approaches [5–8] have also been quite successful and, compared to holistic approaches, are less sensitive to facial expressions, variations in illumination and occlusion. Some of the hybrid approaches include the modular eigenface method [9], the Flexible Appearance Model method [10], and a method that combines component-based recognition with 3D morphable models [11]. Experiments with hybrid methods showed slight improvements over holistic and feature-based methods.

In this work, we evaluated the performance of virtual generalizing random access memory weightless neural networks (VG-RAM WNN [12]) on face recognition using the AR Face Database [13]. There are many face databases freely available for research purposes (see [1] for a comprehensive list); we chose the AR Face Database because we were interested in face recognition under different facial expressions, types of occlusion and illumination conditions, and this database has face images of the same people encompassing all these variations. We evaluated two VG-RAM WNN architectures, one holistic and the other feature-based, each implemented with different numbers of neurons and synapses per neuron. We compared the best VG-RAM WNN performance with that of: (i) a holistic method based on principal component analysis (PCA) [2]; (ii) feature-based methods based on non-negative matrix factorization (NMF) [5], local non-negative matrix factorization (LNMF) [6], and line edge maps (LEM) [7]; and (iii) hybrid methods based on weighted eigenspace representation (WER) [9] and attributed relational graph (ARG) matching [8]. We selected these for comparison because they are representative of some of the best methods for face recognition present in the literature. Our results showed that, even training with a single face image per person, VG-RAM WNN outperformed all mentioned techniques under all face conditions tested.

This paper is organized as follows. Section 2 introduces VG-RAM WNN and Section 3 describes how we have used them for face recognition. Section 4 presents our experimental methodology and experimental results. Our conclusions and directions for future work follow in Section 5.

2 VG-RAM WNN

RAM-based neural networks, also known as n-tuple classifiers or weightless neural networks, do not store knowledge in their connections but in Random Access Memories (RAM) inside the network's nodes, or neurons. These neurons operate with binary input values and use RAM as lookup tables: the synapses of each neuron collect a vector of bits from the network's inputs that is used as the RAM address, and the value stored at this address is the neuron's output. Training can be made in one shot and basically consists of storing the desired output in the address associated with the input vector of the neuron [14].

In spite of their remarkable simplicity, RAM-based neural networks are very effective as pattern recognition tools, offering fast training and test, in addition to easy implementation [12]. However, if the network input is too large, the memory size becomes prohibitive, since it must be equal to 2^n, where n is the input size. Virtual Generalizing RAM (VG-RAM) weightless neural networks (WNN) are RAM-based neural networks that only require memory capacity to store the data related to the training set [15]. In the neurons of these networks, the memory stores the input-output pairs shown during training, instead of only the output. In the test phase, the memory of VG-RAM WNN neurons is searched associatively by comparing the input presented to the network with all inputs in the input-output pairs learned. The output of each VG-RAM WNN neuron is taken from the pair whose input is nearest to the input presented—the distance function employed by VG-RAM WNN neurons is the *hamming distance*. If there is more than one pair at the same minimum distance from the input presented, the neuron's output is chosen randomly among these pairs.

lookup table	$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	**Y**
entry #1	1	1	0	class 1
entry #2	0	0	1	class 2
entry #3	0	1	0	class 3
	↑	↑	↑	↓
input	1	0	1	**class 2**

Fig. 1. VG-RAM WNN neuron lookup table.

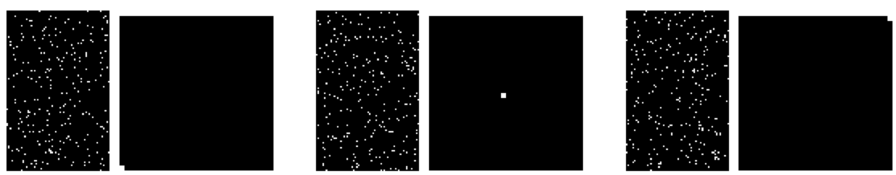
Figure 1 shows the lookup table of a VG-RAM WNN neuron with three synapses (X_1, X_2 and X_3). This lookup table contains three entries (input-output pairs), which were stored during the training phase (entry #1, entry #2 and entry #3). During the test phase, when an input vector (input) is presented to the network, the VG-RAM WNN test algorithm calculates the distance between this input vector and each input of the input-output pairs stored in the lookup table. In the example of Figure 1, the *hamming distance* from the input to entry #1 is two, because both X_2 and X_3 bits do not match the input vector. The distance to entry #2 is one, because X_1 is the only non-matching bit. The distance to entry #3 is three, as the reader may easily verify. Hence, for this input vector, the algorithm evaluates the neuron's output, Y, as class 2, since it is the output value stored in entry #2.

3 Face Recognition with VG-RAM WNN

As stated in the Introduction, in this work we examined the recognition part of the identification problem only. Face segmentation is performed semi-automatically and, thanks to the properties of the VG-RAM WNN architectures employed, explicit feature extraction (e.g., line edge extraction; eye, nose, or mouth segmentation; etc.) is not required, even though in one of the two VG-RAM WNN architectures studied some neurons specializes in specific regions of the faces and, because of that, we say it is feature-based. The other VG-RAM WNN architecture studied is holistic.

3.1 Holistic Architecture

The holistic architecture has a single bidimensional array of $m \times n$ VG-RAM WNN neurons, N, where each neuron, $n_{i,j}$, has a set of synapses $W = \{w_1, \ldots, w_{|W|}\}$, which are randomly connected to the network's bidimensional input, Φ, of $u \times v$ inputs, $\phi_{i,j}$ (Figure 2). This random interconnection pattern is created when the network is built and does not change afterwards.



(a) (b) (c)

Fig. 2. The synaptic interconnection pattern of the holistic architecture. (a) Left, input Φ: in white, the elements $\phi_{i,j}$ of the input Φ that are connected to neuron $n_{0,0}$ of N via $w_1, \ldots, w_{|W|}$; right, neuron array N: in white, the neuron $n_{0,0}$ of N. (b) Left: in white, the elements $\phi_{i,j}$ of Φ connected to $n_{\frac{m}{2},\frac{n}{2}}$; right: in white, the neuron $n_{\frac{m}{2},\frac{n}{2}}$ of N. (c) Left: in white, the elements of Φ connected to $n_{m,n}$; right: in white, the neuron $n_{m,n}$.

VG-RAM WNN synapses can only get a single bit from the input. Thus, in order to allow our VG-RAM WNN to deal with images, we use *minchinton cells* [16]. In the proposed VG-RAM WNN architectures, each neuron's synapse, w_t, forms a minchinton cell with the next, w_{t+1} ($w_{|W|}$ forms a minchinton cell with w_1). The type of the minchinton cell we have used returns 1 if the synapse w_t of the cell is connected to an input element, $\phi_{i,j}$, whose value, $x_{i,j}$, is larger than the value of the element $\phi_{k,l}$, $x_{k,l}$, to which the synapse w_{t+1} is connected (i.e., $x_{i,j} > x_{k,l}$); otherwise, it returns zero.

The input face images, I, of $\xi \times \eta$ pixels (Figure 3(a)), are rotated, scaled and cropped (Figure 3(b)), so that the face in the image fits within the VG-RAM WNN input, Φ. The rotation, scaling and cropping are performed semi-automatically, i.e., the position of the eyes are marked manually and, based on this marking, the face in the image is computationally adjusted to fit into Φ. Before being copied to Φ, the transformed image is filtered by a Gaussian filter to smooth out artifacts produced by the transformations (Figure 3(c)).

(a) (b) (c)

Fig. 3. Face image and its preprocessing. (a) Original image; (b) rotated, scaled and cropped image; and (c) filtered image.

During training, the face image I is transformed and filtered, and its pixels are copied to the VG-RAM WNN's input Φ and all $n_{i,j}$ neurons' outputs, $y_{i,j}$, are set to the value of the label associated with the face (an integer). All neurons are then trained to output this label with this input image. During testing, each face image I is also transformed, filtered, and copied to the VG-RAM WNN's input Φ. After that, all neurons' outputs $y_{i,j}$ are computed and the number of neurons outputting each label is counted. The network's output is the label with the largest count.

3.2 Feature-Based Architecture

As the holistic architecture, the feature-based architecture has a single bidimensional array of $m \times n$ VG-RAM WNN neurons, N, where each neuron, $n_{i,j}$, has a set of synapses, $W = \{w_1, \ldots, w_{|W|}\}$, which are connected to the network's bidimensional input, Φ, of $u \times v$ inputs. The synaptic interconnection pattern of each neuron $n_{i,j}$, $\Omega_{i,j}(W)$, is, however, different (Figure 4). In the feature-based architecture, $\Omega_{i,j}(W)$ follows a bidimensional Normal distribution centered at ϕ_{μ_k,μ_l}, where $\mu_k = \frac{i.u}{m}$ and $\mu_l = \frac{j.v}{n}$, i.e., the coordinates k and l of the elements of Φ to which $n_{i,j}$ connects via W follow the probability density functions:

$$\varphi_{\mu_k,\sigma^2}(k) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(k-\mu_k)^2}{2\sigma^2}} \tag{1}$$

$$\varphi_{\mu_l,\sigma^2}(l) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(l-\mu_l)^2}{2\sigma^2}} \tag{2}$$

where σ is a parameter of the architecture. This interconnection pattern mimics that observed in many classes of biological neurons, and is also created when the network is built and does not change afterwards.

A comparison between Figure 2 and Figure 4 illustrates the difference between the interconnection patterns of the holistic and feature-based architectures. In the feature-based architecture (Figure 4), each neuron $n_{i,j}$ monitors a region of the input Φ and, therefore, specializes in the face features that are mapped to that region. On the other hand, each neuron $n_{i,j}$ of the holistic architecture monitors the whole face (Figure 2).

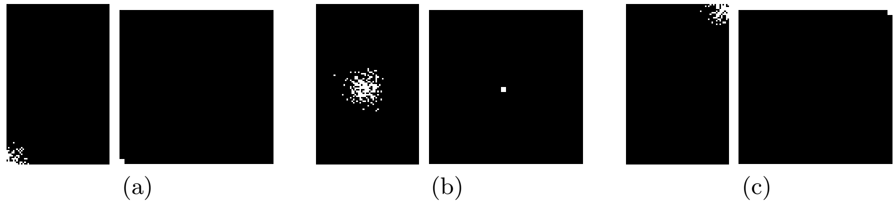


Fig. 4. The synaptic interconnection pattern of the feature-based architecture. (a) Left, input Φ: in white, the elements $\phi_{i,j}$ of the input Φ that are connected to neuron $n_{0,0}$ of N via $w_1, \ldots, w_{|W|}$; right, neuron array N: in white, the neuron $n_{0,0}$ of N. (b) Left: in white, the elements $\phi_{i,j}$ of Φ connected to $n_{\frac{m}{2},\frac{n}{2}}$; right: in white, the neuron $n_{\frac{m}{2},\frac{n}{2}}$ of N. (c) Left: in white, the elements of Φ connected to $n_{m,n}$; right: in white, the neuron $n_{m,n}$.

As in the holistic architecture, in the feature-based architecture each neuron's synapse, w_t, forms a minchinton cell with the next, w_{t+1}, and, before training or testing, the input face images, I, are rotated, scaled, cropped, filtered and only then copied to the VG-RAM WNN input Φ. Training and testing are performed the same way as in the holistic architecture.

4 Experimental Evaluation

We used the AR Face Database [13] to evaluate the performance of VG-RAM WNN on face recognition. This face database contains over 4,000 color images corresponding to 135 people's faces (76 men and 59 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The 768×576 pixels pictures were taken under strictly controlled conditions, but no restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants.

In order to facilitate the comparison with other results in the literature, we used only the following subset of image types of the AR Face Database [13]: neutral expression, smile, anger, scream, left light on, right light on, all side lights on, wearing sun glasses, and wearing scarf. These can be divided into four groups (see Figure 5): (i) normal (neutral expression); (ii) under expression variation (smile, anger, scream); (iii) under illumination changes (left light on, right light on, all side lights on); and (iv) with occlusion (wearing sun glasses, wearing scarf).

We randomly selected 50 people from the database to tune the parameters of the VG-RAM WNN architectures (25 men and 25 women). We used one normal face image of each person to train (50 images), and the smile, anger, wearing sun glasses, and wearing scarf to evaluate the architectures (200 images) while varying their parameters. In the following subsections, we describe the experiments we performed to tune the parameters of the architectures.

(a) (b) (c) (d)

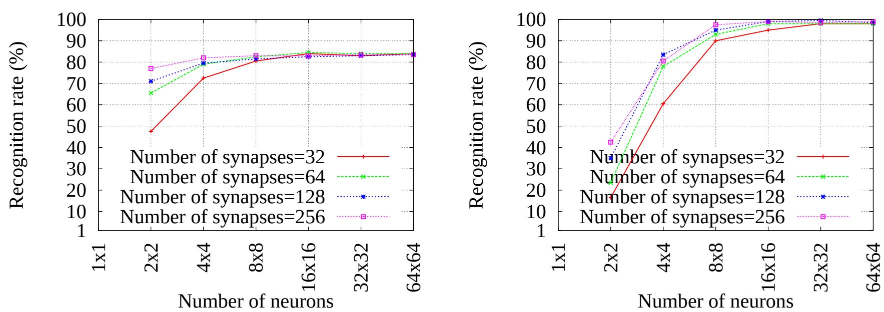
Fig. 5. The AR face database: (a) normal (neutral expression); (b) under expression variation (smile, anger, scream); (c) under illumination changes (left light on, right light on, all side lights on); and (d) with occlusion (wearing sun glasses, wearing scarf).

4.1 Holistic Architecture Parameter Tunning

The holistic architecture has three parameters: (i) the number of neurons, $m \times n$; (ii) the number of synapses per neuron, $|W|$; and (iii) the size of the network input, $u \times v$. We tested networks with: $m \times n$ equal to 2×2, 4×4, 16×16, 32×32 and 64×64; number of synapses per neuron equal to 32, 64, 128 and 256; and $u \times v$ equal to 128×200 (we did not vary $u \times v$ to reduce the parameter search space). Figure 6(a) presents the results of the experiments we carried out to tune the parameters of the holistic architecture. As Figure 6(a) shows, the performance, i.e., the percentage of correctly recognized faces (recognition rate) of the holistic architecture grows with the number of neurons and synapses per neuron; however, as these numbers increase, the gains in performance decrease forming a plateau towards the maximum performance. The simplest configuration in the plateau has around 16×16 neurons and 64 synapses.

4.2 Feature-Based Architecture Parameter Tunning

The feature-based architecture has four parameters: (i) the number of neurons; (ii) the number of synapses per neuron; (iii) the size of the network input; and (iv) σ (see Section 3.2). We tested networks with: $m \times n$ equal to 2×2, 4×4,



(a) (b)

Fig. 6. Performance tunning: (a) holistic architecture and (b) feature-based architecture.

16×16, 32×32 and 64×64; number of synapses per neuron equal to 32, 64, 128 and 256; $u \times v$ equal to 128×200; and σ equal to 10 (we did not vary $u \times v$ and σ to reduce the parameter search space).

Figure 6(b) presents the results of the experiments we conducted to tune the parameters of the feature-based architecture. As Figure 6(b) shows, the performance of the feature-based architecture also grows with the number of neurons and synapses per neuron, and again reaches a plateau at about 32×32 neurons and 128 synapses. However, it is important to note that, in this case, the plateau is very close to 100% accuracy (99.5%).

4.3 Comparison with Other Techniques

We compared the performance of the holistic and feature-based VG-RAM WNN architectures with that of other techniques. For that, we took the best VG-RAM WNN architectures configurations (holistic: 16×16 neurons and 64 synapses per neuron; feature-based: 32×32 neurons and 128 synapses per neuron), trained them with the normal face image of all people in the database (135 images), and tested them with the remaining face images of Figure 5 of all people in the database (135 images of each face image category). Table 1 summarizes this comparison, showing one technique on each line, grouped by type, and the corresponding performance for each face image category on each column.

Table 1. The recognition rate on the AR Face Database. PCA: principal component analysis [2] (results obtained from [8]); VWH: VG-RAM WNN holistic architecture; NMF: non-negative matrix factorization [5] (results from [8]); LNMF: local non-negative matrix factorization [6] (results from [8]); LEM: line edge maps [7] (results from [7] with only 112 people of the AR Face Database); VWF: VG-RAM WNN feature-based architecture; WER: weighted eigenspace representation [9] (results from [9] with only 50 people of the AR Face Database); and ARG: attributed relational graph matching [8] (results from [8]).

Type	Technique	Category							
		Smile	Anger	Scream	Glasses	Scarf	Left light	Right light	All side lights
HOL[a]	PCA	94.1%	79.3%	44.4%	32.9%	2.2%	7.4%	7.4%	2.2%
	VWH	**98.5%**	**97.8%**	**91.1%**	**66.7%**	**25.2%**	**97.8%**	**95.6%**	**95.6%**
FBA[b]	NMF	68.1%	50.4%	18.5%	3.7%	0.7%	N/A[d]	N/A	N/A
	LNMF	94.8%	76.3%	44.4%	18.5%	9.6%	N/A	N/A	N/A
	LEM	78.6%	92.9%	31.3%	N/A	N/A	92.9%	91.1%	74.1%
	VWF	**99.3%**	**99.3%**	**93.3%**	**85.2%**	**98.5%**	**99.3%**	**98.5%**	**99.3%**
HYB[c]	WER	84.0%	94.0%	32.0%	80.0%	82.0%	N/A	N/A	N/A
	ARG	97.8%	96.3%	66.7%	80.7%	85.2%	98.5%	96.3%	91.1%

[a]HOL: holistic techniques. [b]FBA: feature-based techniques. [c]HYB: hybrid techniques.
[d]N/A: not available.

As Table 1 shows, the VG-RAM WNN holistic architecture (VWH) performed better than all holistic and feature-based techniques examined in all face image categories. It also performed better than the hybrid techniques, except for the categories with occlusion and single side illumination. That was expected, since occlusions and single side illumination compromise eventual similarities between the input patterns learned by the VWH neurons and those collected by its synapses from a partially occluded or illuminated face. Nevertheless, it is important to note the overall performance achieved by VWH, which is better than that of several other relevant techniques from literature.

The VG-RAM WNN feature-based architecture (VWF) performed better than all other techniques examined in all categories, in many cases for a significant margin. This performance is the result of two factors. First, each VWF (or VWH) synapse collects the result of a comparison between two pixels, executed by its corresponding minchinton cell. Our best VWF has 128 synapses per neuron and 32×32 neurons. Therefore, during test, 131072 ($128 \times 32 \times 32$) such comparisons are executed on an input face image and the results are checked against equivalent results learned from training images. This amount of pixel comparisons allows not only high discrimination capability but also generalization. Second, thanks to the characteristics of the VWF architecture, i.e., its synaptic interconnection pattern, each VWF neuron monitors a specific region of the face only, which reduces the overall impact of occlusions and varying illumination conditions on recognition performance.

5 Conclusions and Future Work

In this work, we presented an experimental evaluation of the performance of Virtual Generalizing Random Access Memory Weightless Neural Networks (VG-RAM WNN [12]) on face recognition. We presented two VG-RAM WNN face recognition architectures, one holistic and the other feature-based, and examined its performance with a well known face database, the AR Face Database. This database is challenging for face recognition systems because it has images with different facial expressions, occlusions, and varying illumination conditions. The best performing architecture (feature-based) showed robustness in all image conditions and better performance than many other techniques from literature, even when trained with a single sample per person.

In future works, we will examine the performance of VG-RAM WNN with other databases and use it to tackle other problems associated with face recognition systems, such as face detection, face alignment, face recognition in video, etc.

6 Acknowledgements

We would like to thank *Receita Federal do Brasil*, *Conselho Nacional de Desenvolvimento Científico e Tecnológico* — CNPq-Brasil (grants 308207/2004-1, 471898/2004-0, 620165/2006-5, 309831/2007-5), *Financiadora de Estudos e Projetos* — FINEP-Brasil (grants CT-INFRA-PRO-UFES/2005, CT-INFRA-PRO-UFES/2006), and *Fundação Espírito Santense de Tecnologia* — FAPES-Brasil (grant 37711393/2007) for their support to this research work.

References

1. Zhao, W.Y., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys **35**(4) (2003) 399–458
2. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **3** (1991) 71–86
3. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7) (1997) 711–720
4. Etemad, K., Chellappa, R.: Discriminant analysis for recognition of human face images. Journal of the Optical Society of America A **14**(8) (1997) 1724–1733
5. Lee, D.D., Seung, S.H.: Learning the parts of objects by non-negative matrix factorization. Nature **401** (October 1999) 788–791
6. Li, S.Z., Hou, X.W., Zhang, H., Cheng, Q.: Learning spatially localized, parts-based representation. In: Proceedings of Computer Vision and Pattern Recognition. (December 2001) 207–212
7. Gao, Y., Leung, M.K.: Face recognition using line edge map. IEEE Transactions Pattern Analysis and Machine Learning **24**(6) (June 2002) 764–779
8. Park, B.G.: Face recognition using face-arg matching. IEEE Trans. Pattern Anal. Mach. Intell. **27**(12) (2005) 1982–1988
9. Martínez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(6) (June 2002) 748–763
10. Lanitis, A., Taylor, C.J., Cootes, T.F.: Automatic face identification system using flexible appearance models. Image Vision Computing **13**(5) (1995) 393–401
11. Huang, J., Heisele, B., Blanz, V.: Component based face recognition with 3d morphable models. In: Proceedings of the International Conference on Audio- and Video-Based Person Authentication. (2003) 27–34
12. Aleksander, I.: From WISARD to MAGNUS: a Family of Weightless Virtual Neural Machines. In: RAM-Based Neural Networks. World Scientific (1998) 18–30
13. Martinez, A., Benavente, R.: The AR face database. Technical Report #24, Computer Vision Center (CVC), Universitat Autònoma de Barcelona (1998)
14. Aleksander, I.: Self-adaptive universal logic circuits. IEE Electronic Letters **2**(8) (1966) 231–232
15. Ludermir, T.B., Carvalho, A.C.P.L.F., Braga, A.P., Souto, M.D.: Weightless neural models: a review of current and past works. Neural Computing Surveys **2** (1999) 41–61
16. Mitchell, R.J., Bishop, J.M., Box, S.K., Hawker, J.F.: Comparison of Some Methods for Processing Grey Level Data in Weightless Networks. In: RAM-Based Neural Networks. World Scientific (1998) 61–70