Memory Efficient Sign Language Recognition System Based on WiSARD Weightless Neural Network technique

Faisal Ghazanfar[1],Najmi Ghani Haider[2]

[1]Applied physics Computer and instrumentation center
PCSIR Labs complex KarachiKarachi, Pakistan

faisalku2000@gmail.com

[2]Department of Computer Science & Information Technology,
NED University of Engineering and Technology, Karachi, Pakistan

najmi@neduet.edu.pk

Abstract: A novel single hand Sign Language Recognition (SLR) method utilizing Weightless Neural Network (WNN) known as a RAMnet or n-tuple network is evaluated as a memory efficient technique. In contrast with standard multilayer perceptron neural network (MLPNN), the RAMnet does not require long iteration of presentation of training data in the training phase i.e. long training time, weights adjustments and activation function calculations. This alternative technique is based on WiSARD Neural Network. Two algorithms are proposed and implemented: one for hand tracking through image sequences, and the second for SLR using WiSARD. The pixels containing the hand region are used to train the Sign Language (SL) alphabetical symbol discriminators. Each sign is assigned a discriminator, thus there are as many discriminator as signs to be recognized. The system performance is checked by evaluating for single handed SL English alphabets and produced result of accuracy 75%. The conventional Random Access Memory (RAM) is used to save and process binary image sequences in WiSARD that are addressed by Boolean input and produced Boolean output. The system can implement signs for any other sign language, thus it is language independent.

Keywords and phrases: Recognition, Sign Language (SL), MLPNN, WiSARD.

I. BACKGROUND AND INTRODUCTION

Visual gesture is a very common communication medium for persons with hearing disability. Societies of hearing disabled persons around the world have developed their own gestures according to the requirements of their society. There are 25,000 signs in deaf people's sign language that are in communication but only a small percentage of these have been recorded in the several signs languages dictionaries published [5,6]. The fuzzy logic and Artificial Neural Network (ANN) models are commonly applied for hand recognition with in some shortcomings. Fuzzy Logic (FL) is based on probabilities for different possibilities. It consists of many statistical calculations such as standard deviation, variance, covariance, Eigenvectors and Eigen values. The colored glove approach for the recognition of Arabic sign language is a good example that avoided the segmentation problem and obtained unique features [12]. In image recognition such multifaceted processing, not only increases complexity but also needs high processing time.

ANN is based on weighted-sum-and-threshold artificial neurons. It requires tedious empirical evaluation to establish neural network architecture. Training requires propagation in forward as well as in backward phase so complexity increases when the input data is large in size.

The Random Access Memory Weightless Neural Network (RAM-NN) is an efficient machine learning technique that provided the simplest training and testing capability[13]. RAM size is determined bytuple size*n*, i.e. 2^nas shown in table 1 (for *n*= 4).

TABLE 1
Relationship of image and RAM size (n=4)

Image size	No. of Pixels (A)	No. of Tuples T= A/4	RAM required 2^nx T (bits)	Total RAM required (bytes)
16x16	256	64	2^4x 64	128
320x240	76,800	19,200	2^4 x 19200	3.838K
800x600	480,000	12,0000	2^4 x 120,000	240K

The RAM-NN is a weightless Neural Networks (WNN) model which requires low memory capacity to save the data associated to training [2]. This technique stores pixel positions as a

978-1-5090-4059-9/16/$31.00 ©2016 IEEE

number of tuple at a specific address in memory. Every Training image generates a partial pattern of pixels with its unique address. Once training is completed, the RAM-NN is ready for testing the trained information (recognition), it searches the associative neuron by comparing the input given to the network with all learned input-output to predict the correct outcome.

Due to such characteristics of WNN it has been used as a highly simplified learning technique and an excellent pattern classifier. The fast processing, easy learning scheme and readability of WNN programs enhances its capability and implementation on dedicated hardware.

Bledsoe and Browning [9] suggested WNN as a pattern recognition method where learning to recognize an image can be through a set logical functions that can describe the problem. WNN do not save information in their connections, it uses binary values as input vector and utilized the RAM of computer as a lookup table; each neuron collects bits information from the network that is used as address of RAM, and 0's and 1's saved at this address is worked as output of the neuron [10].

The problem statement of this research is formulated as (i) training phase, (ii) testing phase. The system takes the input image and extracts the hand region for identification; the system confirms or rejects the hand gesture based on its training.

II. THE WiSARD APPROACH

The single hand sign language recognition system is utilizing a recognition device known as WiSARD (Wilkie, Stonham and Aleksander's Recognition Device) [1]. WiSARD uses binary image as the input pattern, processes it and generates a score for that pattern. The Discriminators are used for learning and recognition purpose; a summing device with n-inputs and a set of X one-bit word RAM are the major components of WiSARD. All functions of WiSARD are carried out by RAM-discriminators. A binary pattern is given as input to each RAM-discriminator that produces a response against every input. The number of responses is evaluated by a training algorithm, which compares them and computes the relative result of the highest response or highest score as shown in figure 1.

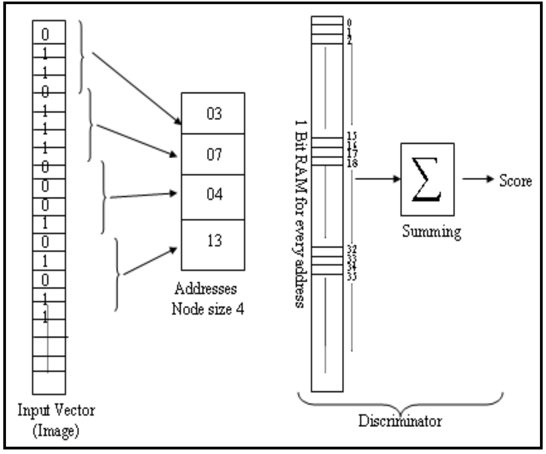

Figure 1: The binary input associated to the addresses, one square box is one discriminator of size 24 bits in RAM of computer (Discriminator). The score is generated by the sum of all mapped 1s and 0s of every discriminator (Boolean Model -WiSARD)

The score of a discriminator is a characteristic property of that mental image or input. It is dependent on the group of pixels (called tuple), if the size of tuple is 4 its means there are 16 addresses in a 1-bit RAM cell, while if the tuple size is 6 it is 64as given in table 1. Therefore the tuple size and RAM's address are related to power of 2 [9].

A. WiSARD and Image Processing

The WiSARD model is a different to standard image processing (IP). The image is not processed pixel by pixel [2,3,4], but only by 1-bit RAM blocks in discriminator is required to process the image. Since it is a supervised learning approach, therefore the training compensates the noise and environmental errors through threshold setting in binarization, while in IP these parameters are a preprocessing task and a separate algorithm is needed to deal with it[1].

The WiSARD is a device which stores 1-bit information for every address. The input is controlled in such a way that the RAM is set for either "Write" or "Read". The writing phase is a "Learning phase" while the reading phase is "Test/operation phase". Initially all discriminator (RAMs) are cleared (i.e. set to zero). During the Training phase the memory set to '1' for every address produced according to the *n*-tuple; In testing phase (reading mode) the output returns back to the corresponding address of tuples for the trained pattern. RAM's addressed are read and the output mapped to produce result as shown in figure 2.

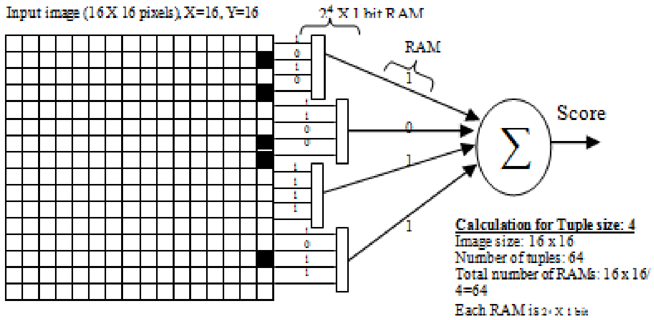
Input image (16 X 16 pixels), X=16, Y=16
2^4 X 1 bit RAM
RAM
Score
Σ
1
0
1
1
Calculation for Tuple size: 4
Image size: 16 x 16
Number of tuples: 64
Total number of RAMs: 16 x 16/ 4=64
Each RAM is 2^4 X 1 bit

Figure 2: Mapping of image to RAM to produced result score

III. SIGN LANGUAGE RECOGNITION SYSTEM

The sign language recognition system is based on learning regarding the alphabetical signs and its verification or testing [7]. The processes have been done through two algorithms; one for training and other for testing. The system is also accomplished in two phases; Phase 1: The first phase is the training phase where the system finds the area of processing in the image that is the area where hand is exists. Phase 2: The second phase is recognition phase where the sign alphabet is recognized by computing the highest scoring discriminator and its associated sign.

The algorithms for hand gesture and sign language recognition are developed and implemented in MatLab that are utilized for camera interfacing and configuration to access every 5th frame, as well as binarization of color frames for recorded movie and snapshots for real time recognition.

A .Training Phase

The Training/Learning phase is worked in two areas; background learning, and extraction of hand from snapshot/frame i.e. finding the area of interest. Figure 3gives a summary of background training where a person sits in front of a camera on a chair with hands down (i.e. no hand showing in the image) for a few seconds (the frames captured are called Mental images). The training completes when the desired confidence level is achieved (set as a percentage by the user) where confidence level is the results' difference between the current and previous scores divided by the maximum score.

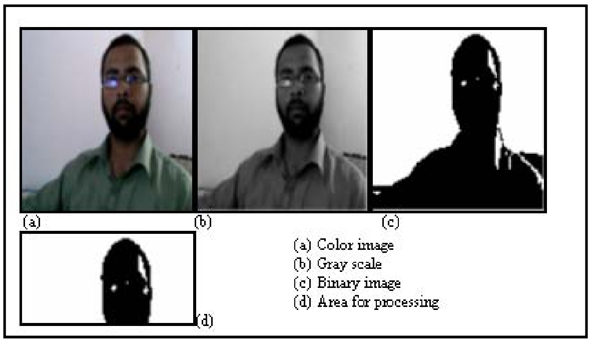

Figure3: Stages towards training phase, reduced size of snapshot used to search hand.

The size of image acquired through camera is 120x160 while the processing size in this experiment is 100x50 and tuple size selected for WiSARD is 4. The relationship between size of tuple and RAM size is given as

$$N = \quad 2^S \qquad (1)$$

Where *S* is the size of tuple and *N* is the size of RAM, since tuple size is 4 therefore

$$N = 2^4 = 16$$

Each tuple will be mapped to 16 x 1-bit RAM, therefore the total number of Tuples (T) are

$$T = \frac{Number\ of\ pixels}{Tuple\ size} \qquad (2)$$

$$T = \frac{100 \times 50}{4} = 1250$$

Hence RAM used to process a 100x50 size image is 1250x16x1 bits or 2.5 Kbytes. Low memory consumption enables WiSARD to process an image rapidly.

a) Background training

A common simple web camera is attached to the computer with 30 fps and triggers every 5th frame for analysis. The frame is captured through camera and converted into gray scale that is all pixel values in the range from 0 to 255. This gray image is then converted into binary image using global threshold method. The goal of binarization is to differentiate features from background so that counting, matching and measurement operations can be performed. The binarization reduced the complexity of the process as shown in figure 3.

The binary image pattern is mapped in a sequence of numbers of pixels according to size of image. Total 100 x 50 = 5000 pixels are mapped to an array of size 5000 x 2. *S* numbers of pixels are read from mapped array and converted into decimal form to produce address saved in tuple address array. In WISARD the address is a physical location of conventional Random Access Memory (RAM) having size 1 bit per tuple. For every tuple there are 16 locations in RAM that are written by ‘1’ according to address generated by the tuple.

The total numbers of 1’s in RAM are summed up to get the Result Score (R). The Training score is changed for every frame of input so it can be monitored by difference of frame from previous frame. The training is complete when the difference of frames’ scores is achieves the set confidence level. Since we are interested to recognize hand for SLR so the upper half area of image is utilized for the training i.e. it is the place in the image where the hand will be appear.

b) Hand Finding in Image

The operation of WiSARD uses negative logic method for searching hand in the image. The background training data does not contain hand information therefore when the image sequence with hand is brought into view, the discriminator output changes with respect to previous frame of image as well as training data. This change of discriminator output is utilized to indicate the presence of the hand in the image. Furthermore, the location in the image where the hand is to be found is obtained by finding the changed number of 1’s in discriminator (saved in RAM). Therefore the method for removing background and searching of hand in image is generic.

The Hand finding algorithm uses a single discriminator trained on background image. The overall process is that a person sits on a chair with his hand not appearing in the image (camera is showing face and shoulders). This setup is for hand background learning. When this background training is completed, the system can now be used in operational mode for hand detection. The step in algorithm for hand tracking task is given in figure4.

The location of hand is determined either at right side of the head or the left side by looking at RAM addresses of these two areas (only for upper half portion of image). The binary input image with hand is applied to the discriminator of WiSARD model. It compares it to the corresponding background trained data sets and generates a left counter and right counter that determines the presence of hand in the image(figure 5).

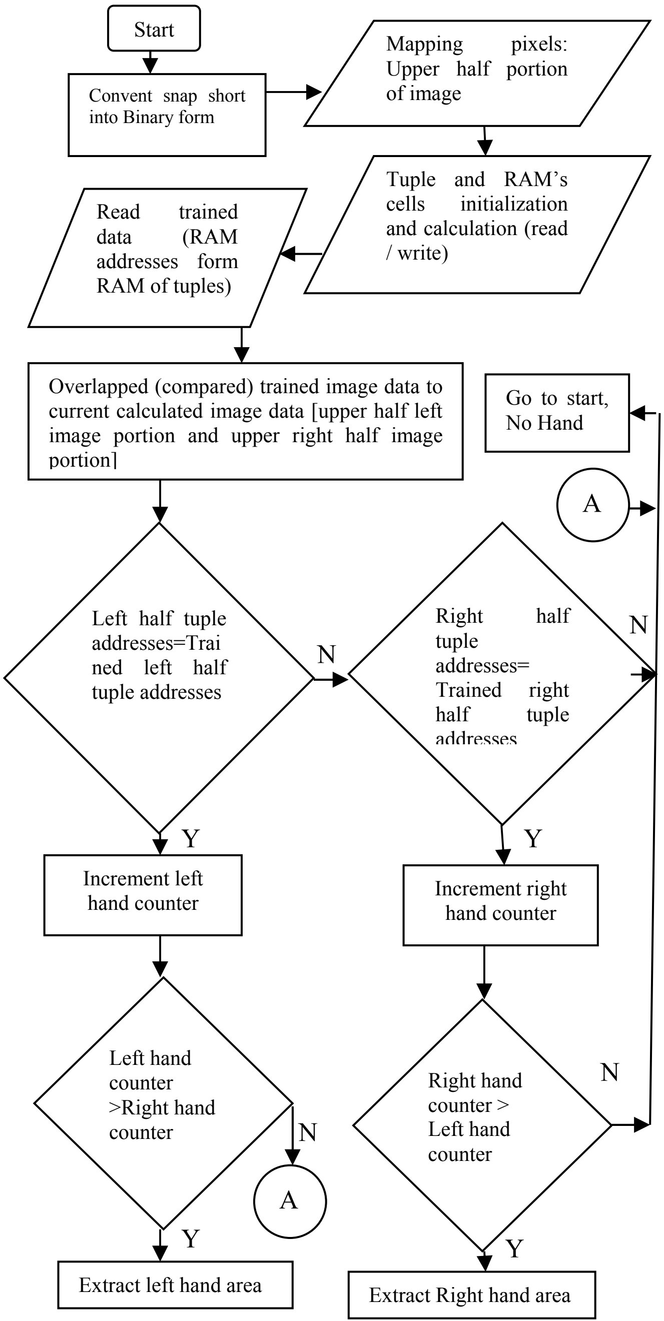
Start
Convent snap short into Binary form
Mapping pixels: Upper half portion of image
Tuple and RAM’s cells initialization and calculation (read / write)
Read trained data (RAM addresses form RAM of tuples)
Overlapped (compared) trained image data to current calculated image data [upper half left image portion and upper right half image portion]
Go to start, No Hand
A
Left half tuple addresses=Trained left half tuple addresses
N
Right half tuple addresses= Trained right half tuple addresses
N
Y
Y
Increment left hand counter
Increment right hand counter
Left hand counter >Right hand counter
N
A
Right hand counter > Left hand counter
N
Y
Y
Extract left hand area
Extract Right hand area

Figure4: The flow chart for hand finding algorithm from Image is a part of background training

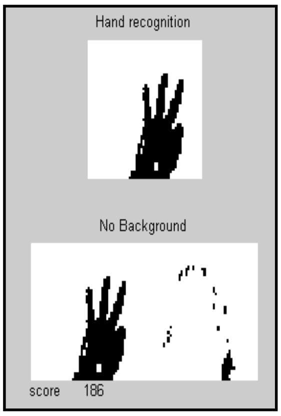

Figure 5: (a) The right hand from snapshot is extracted during training (upper part of image)
(b) Image processing of selected region (Lower part of the image)

IV. SIGN RECOGNITION

The goal of Sign Language recognition using WiSARD is the development and testing of algorithms that can be easily implemented on a portable hardware such as a mobile phone [1, 8]. This model can resolve the problems in portable devices where we have restrictions of memory, limitation in speed and processing. The discriminator associated with hand detection contains information about hand presence in the image (training phase). When the hand is brought in view of camera, it locates the hand in the image, i.e. the image area containing the hand, by detecting abrupt change in the trained discriminator output. The process of training is repeated for each gesture therefore the number of discriminators is as many as the number of signs, as shown in figure 6.

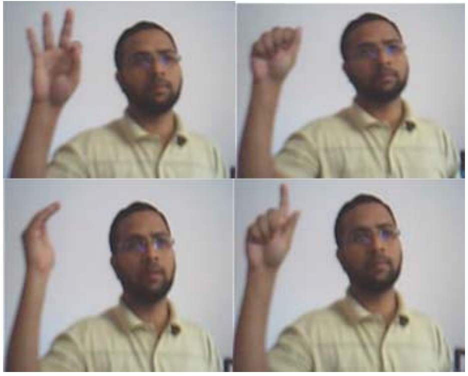

Figure6:Snapshots/Frames for Symbol A, B, C and D (top left, top right, bottom left, bottom right)

In the overall process the main program performed matching of input image pattern to every sign trained discriminator and produced output as a score as shown in figure 7. The discriminator producing the highest score is selected and the associated hand sign is output. Since a small amount of memory is processed for recognition of a given pattern of image, the technique is very quick and is easily implementable in hardware. Every hand sign has its own discriminator, so the number of discriminators is equal to the number of hand signs. The hand sign presented to the camera produces a score from every discriminator with the highest scorer recognized as the correct sign. Since all hand signs are also trained in the same manner as background, so body movement or shaking issue is also there, therefore the average of score for signs are used for recognition instead of a single value to smooth out these effects.

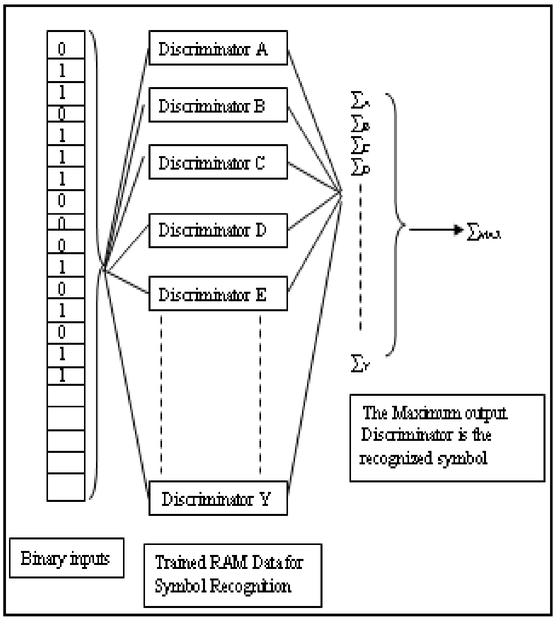

Figure7: Sign Recognition System. The response of every discriminator observed for a given input image.

V. OBSERVATIONS AND RESULTS

Sign language recognition in this paper is initially designated for static gestures only; therefore the results of dynamic gestures\symbols are not included, such as symbol for 'Z' and 'J'. Furthermore a few signs are also not included in recognition due to similarity in outputs, i.e. M and N, E and B. These are called ambiguities and open problem.

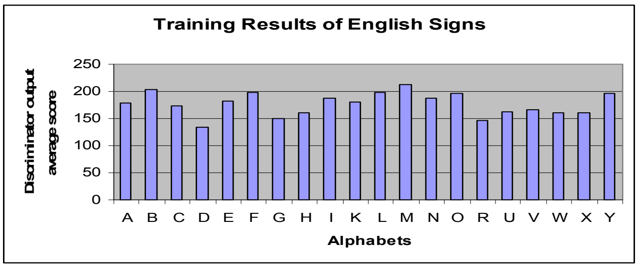

Figure 7: Average discriminator output of symbols

The results for English alphabets sign recognition is shown in figure 7. Fifteen tests were performed for each static symbols in the sign language; therefore the total number of observations obtained is 300. The accuracy for every symbol is calculated by number of symbols input to the system and number of symbols recognized by the system as given in table 2.

$$Accuracy = \frac{symbols\ viewed\ to\ the\ system}{symbols\ recognized} \times 100 \qquad (3)$$

The Evaluation of WiSARD provided symbol recognition accuracy of Sign language. The 'Z' hand sign is not included in the test due to its dynamic nature of symbol. Although a few symbols are not recognized fully, 12 out of the 21 symbols were recognized with 100% accuracy, with average accuracy for the single hand sign language recognition of 75%.

TABLE 2:
Test results of static English alphabets signs

Symbol/sign	No. of Tests (a)	Test passes (b)	Accuracy (b/a*100)%
A	15	6	40
B	15	3	20
C	15	15	100
D	15	15	100
E	15	3	20
F	15	15	100
G	15	15	100
H	15	6	40
I	15	15	100
K	15	6	40
L	15	15	100
M	15	3	20
N	15	6	40
O	15	15	100
P	15	6	40
R	15	6	40
U	15	15	100
V	15	15	100
W	15	15	100
X	15	15	100
Y	15	15	100

Average Accuracy:75%

A. Open Problems

The low memory requirement in WiSARD model enables imaging applications to be developed and implemented on a mobile phone [1]. The Java Virtual Machine, Net beans, Nokia SDK and Nokia PC suite are essential for implementation.

This research used linear mapping to n-tuples. Based on the experience of other researchers [11] it is expected that a scheme using random mapping of pixels to tuples will improve accuracy.

B. Ambiguities

Some ambiguities such as dynamic gestures (symbol of Z) can be resolved by using two cameras. The cameras would be mounted in different directions such that finger movements can be viewed and hence easier to recognize. Hybrid neural network models can be developed process signs involving hand movement.

VI. CONCLUSION

A weightless neural network WiSARD model is proposed and evaluated for static single hand sign language recognition. Conventional RAM of size 2^n x 1 bits can be utilized to process image frames, thus the model provides fast processing and small memory storage, hence realizable hardware implementation in real time environment. The design has produced promising results that need to be improved further for an embedded system.

A novel method for fast hand tracking as well as sign language recognition is developed and demonstrated, without pixel processing in software. This makes the design implementable in real time on portable devices such as a smart phone. The paper discussed proof of concept, and further improvement in performance should be achievable, e.g. through random mapping of tuples.

REFERENCES

[1] E.do.VSimões, L F Uebelet al., “*Hardware Implementation of RAM Neural Networks”,* Informatics Institute – P R Letters, Volume 17, Issue 4, Pages 421-429 ,4 April 1996.

[2] X. Zabulisy, H. Baltzakisy, et al., “*Vision- based Hand Gesture Recognition for Human-Computer Interaction*”, Institute of Computer Science Foundation for Research and Technology - Greece, IEEE Trans. on Pattern Analysis and Machine, 1997.

[3] C. Manresa, J.Varona et al.; “*Real –Time Hand Tracking and Gesture Recognition for Human-Computer Interaction”,* E. Letters on Computer Vision and Image Analysis; 2000.

[4] M. E. Petersena, D. de Ridderbet al. *Image processing with neural networks—a Review*; J. of Pattern Recognitionpage35, 2002.

[5] K. Assaleh and M. Al-Rousan, “*Recognition of Arabic Sign Language Alphabet Using Polynomia Classifiers”,* J. on Applied Signal Processing , Volume 2005

[6] A.K. Alvi andM.Y.Butt, *“Pakistan Sign Language Recognition Using Statistical matching”* World academy of science, Engineering and Technology, volume 3, 2005

[7] S. B. Kotsiantis, “*Supervised Machine Learning: A Review ofClassification Techniques”*, Informatica-31, pg. 149-268, 2007

[8] G.Caridakis, O.Diamantiet al. “*Automatic sign language recognition: Vision based feature extraction and probabilistic recognition scheme from multiple Cues”,* PETRA-08, USA

[9] I. Aleksander, M. De Gregorio, F.M.G. França, P.M.V. Lima, H. Morton; *A brief introduction to Weightless Neural Systems;* ESANN'2009. Proceedings, European Symposium on Artificial Neural Networks – Advances in Computational Intelligence and Learning. Bruges(Belgium), 22-24 April 2009.

[10] I. Aleksander, et al.,RAM-Based Neural Networks chapter “*From WISARD to MAGNUS: a Family of Weightless Virtual Neural Machines*”, , World Scientific, pp. 18–30,1998.

[11] R. J. Mitchell, J.M. Bishop, S.K. Box, and J.F. Hawker, RAM-Based Neural Networks chapter “*Comparison of Some Methods for Processing Grey Level Data in Weightless Networks*”, World Scientific, pp. 61–70,1998

[12] M. De Gregorio and M. Giordano, “*Change Detection with Weightless Neural Networks*” IEEE Conference on Computer Vision and Pattern Recognition Workshops 23-28 June 2014, Columbus, Ohio.

[13] F.M.G. França, M. De Gregorio, P.M.V. Lima, W.R. de Oliveira, “*Advances in weightless neural Networks*” ESANN 2014