Stereo Matching with VG-RAM Weightless Neural Networks

Lucas de Paula Veronese, Lauro José Lyrio Junior, Filipe Wall Mutz, Jorcy de Oliveira Neto, Vitor Barbirato Azevedo, Mariella Berger, Alberto Ferreira De Souza and Claudine Badue
Departamento de Informática
Universidade Federal do Espírito Santo
Vitória-ES, Brazil
lucas.veronese@lcad.inf.ufes.br

***Abstract*—Virtual Generalizing Random Access Memory Weightless Neural Networks (VG-RAM WNN) is an effective machine learning technique that offers simple implementation and fast training and test. We examined the performance of VG-RAM WNN on binocular dense stereo matching using the Middlebury Stereo Datasets. Our experimental results showed that, even without tackling occlusions and discontinuities in the stereo image pairs examined, our VG-RAM WNN architecture for stereo matching was able to rank at 114th position in the Middlebury Stereo Evaluation system. This result is promising, because the difference in performance among approaches ranked in distinct positions is very small.**

Keywords-component; Binocular Dense Stereo Matching, VG-RAM Weightless Neural Networks, Middlebury Stereo Vision Page

I. INTRODUCTION

The images projected inside our eyes are constantly changing due to the movement of the eyes or the body as a whole. However, in an apparent paradox, we perceive the world, depicted in the images captured by the eyes, as stable. Moreover, the images projected on human retinas are two-dimensional; however, the brain is able to synthesize a stable three-dimensional representation from them (what we see, Figure 1), with color, shape and depth information about the objects in the surrounding environment, eliminating the effects of the eyes and body movements.

The biological visual system enables our movement through the three-dimensional environment accurately.So, the understanding and modeling of relevant capabilities of the biological visual system, such as those that allow us to see in three dimensions, may contribute to the development of Simultaneous Localization And Mapping (SLAM) [1]and navigation systems for autonomous vehicles.

SLAM is perhaps the most fundamental problem of autonomous robotics. Autonomous vehicles need to know where they are in their area of operation and how this (area of operation) is configured so that they can navigate and perform their activities of interest. For this, autonomous vehicles must use sensors for acquiring sufficient information for mapping the environment and localizing themselves in the generated map. Currently, one of the most widely used sensors areLaser Range Scan, or Light Detection And Ranging (LIDAR).

LIDAR sensors employ an opto-mechanical scanning that makes use of laser beams to measure the distance along straight lines to obstacle points ahead of the vehicle. Such a mechanism is not biologically plausible. Also, LIDAR sensors are strongly affected by rain, among other weather conditions, which limits their applicability in scenarios that require outdoor operation. Moreover, they are easily detectable from a distance (because of the laser), which limits their military applicability.

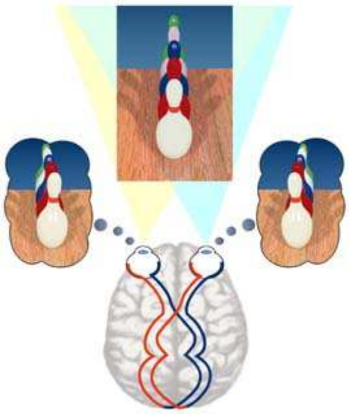

Figure 1: Generation, by the brain, of a three-dimensional representation from two-dimensional images.

Digital cameras, on the other hand, are now able to capture images with millions of pixels and have a substantially lower cost than LIDAR sensors. However, in order to employ digital cameras for solving the SLAM problem in real time, it is necessary to process the huge amount of data captured by the cameras in a manner equivalent to that in our brain, i.e., it is necessary to synthesize stable three-dimensional representations from two-dimensional images.

To generate a three-dimensional representation from two-dimensional images, it is necessary to localize corresponding pixels in images captured by two or more cameras positioned in distinct spatial locations. This problem is known as the *stereo matching problem*. Using the information about the location of corresponding pixels in the several images, and the knowledge of the geometry and positioning of each camera, it is possible to solve the problem of perceiving the world in 3D based on images captured by two or more cameras.

The state-of-the-art stereo matching algorithm is the AD-Census [3]. It calculates the dense disparity map in two main steps. In the first step, it initializes the disparities using the AD-Census metric. In the second step, it uses a cross based cost aggregation algorithm [2] to reduce ambiguities in textureless areas. Other approaches have been proposed to tackle the problem of stereo matching, such as those based on belief propagation [4, 5], self organizing neural networks [6] and disparity energy models [7].

In our search to understand the brain, we developed successful biologically inspired applications, for example, face recognition [8], text categorization [9] and depth estimation using monocular cues [10].Another important aspect of the human perception of the world is the depth perception via stereopsis [11]. To emulate this capacity in an artificial system, we use a stereo camera and a biologically inspired stereo matching algorithm. In fact, in all of these applications, we use Virtual Generalizing Random Access Memory Weightless Neural Networks (VG-RAM WNN) [12]. The VG-RAM WNN is an effective machine learning technique that offers simple implementation and fast training and test.

In this paper, we evaluate the performance of VG-RAM WNN on stereo matching. Our approach addresses the problem of binocular stereo dense matching, i.e., it operates on two images under known camera geometry and computes a dense disparity map that contains a disparity estimate for each pair of corresponding pixels in the two images.

To evaluate the performance of VG-RAM WNN on stereo matching, we used the Middlebury Stereo Datasets (http://vision.middlebury.edu/stereo/data/) [13, 14]. We chose these datasets because we were interested in comparing our experimental resultswith those submitted to the Middlebury Stereo Evaluation system (http://vision.middlebury.edu/stereo/eval/). Our experimental results showed that, even without tackling occlusions and discontinuities in the stereo image pairs examined, our VG-RAM WNN architecture for stereo matching was able to rank at 114th position in the Middlebury Stereo Evaluation system.

This paper is organized as follows. After this introduction, in Section II, we introduce VG-RAM WNN and, in Section III, we describe how we have used them for stereo matching. In Section IV, we describe our experimental methodology and, in Section V, we analyze our experimental results. Our conclusions and directions for future work follow in Section VI.

II. 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 [15].

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 [16]. In the neurons of these networks, the memory stores the input-output pairs shown during training, instead of 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.

Table 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 Table 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, becauseX_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 *output 2*, since it is the output value stored in *entry #2*.

Table 1: VG-RAM WNN neuron lookup table.

Lookup table	X_1	X_2	X_3	Y
entry #1	1	1	0	*output 1*
entry #2	0	0	1	*output 2*
entry #3	0	1	0	*output 3*
	↑	↑	↑	↓
input	1	0	1	***output 2***

III. STEREO MATCHING WITH VG-RAM WNN

Stereo matching is the problem of localizing corresponding pixels in multiple images of the same 3D view captured by cameras in distinct spatial locations. In most camera configurations, finding correspondences requires a search in multiple dimensions.However, if the cameras are aligned to be epipolar using an image rectification algorithm [17], the search for corresponding pixels is simplified to one dimension (a straight line parallel to the baseline between the cameras), i.e., stereo matching can be made in a single scan line. In this paper, we assume that the epipolarity constraint is guaranteed.

Our VG-RAM WNN architecture for stereo matching has a single two-dimensional array of $m \times n$ neurons,N, where each neuron,$n_{i,j}$, has a set of synapses,$W = (w_1, w_2, \ldots w_{|W|})$, which are connected to the network's two-dimensional input, Φ, of $m \times n$ inputs, $\varphi_{i,j}$ (Figure 2). (Note that the network input, Φ, has the same size of the neurons array, N.) The synaptic interconnection pattern of each neuron $n_{i,j}$, $\Omega_{i,j,\sigma}(W)$, follows a two-dimensional Normal distribution with variance σ^2centered at$\varphi_{i,j}$; 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:

$$\omega_{i,\sigma}(k) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{(k-i)^2}{2\sigma^2}}$$

$$\omega_{j,\sigma}(l) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{(l-j)^2}{2\sigma^2}}$$

where$\sigma$is a parameter of the architecture. This synaptic interconnection pattern mimics what is observed in many classes of biological neurons [18], and is created when the network is built and does not change afterwards.

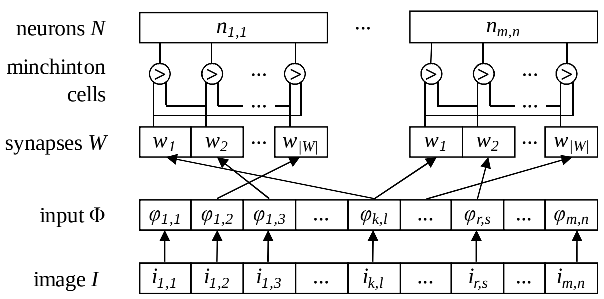

Figure 2: Schematic diagram of our VG-RAM WNN architecture for stereo matching.

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, in which a pixel may assume a range of different values, we use *minchinton cells*[19]. In the proposed VG-RAM WNN architecture, 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, $\varphi_{k,l}$, whose value is larger than the value of the element $\varphi_{r,s}$to which the synapse w_{t+1} is connected, i.e., $\varphi_{k,l} > \varphi_{r,s}$; otherwise, it returns zero (see the synapses w_1 and w_2 of the neuron $n_{m,n}$of Figure 2).

To compute the disparity map between two stereo images, the left image, I_L, is used to generate the training set and the right image, I_R, is used to generate the test set.(A disparity map is an image where each pixel corresponds to the distance in pixels between corresponding pixels in the left and right images of a stereo image pair [17].).

During the training phase, the pixels of I_L are copied to the VG-RAM WNN's input Φ. In the first iteration, the disparity value, d, is set to zero and all $n_{i,j}$ neurons' outputs are also set to d=0. All neurons are then trained to output d=0. Remember that the synaptic interconnection of each neuron $n_{i,j}$ is centered at $\varphi_{i,j}$). In the second iteration, d is incremented by one (d=1) and the array of neurons is shifted of one column to the right, such that the synaptic interconnection of each neuron $n_{i,j}$ is now centered at $\varphi_{i,j+d} = \varphi_{i,j+1}$). All neurons are then trained to output d=1 (except those in column 1). This process iterates until d reaches a maximum pre-defined disparity value, d_{max} (d=d_{max}). In this last iteration, the array of neurons is shifted of d_{max} columns to the right, such that the synaptic interconnection of each neuron $n_{i,j}$ is centered at $\varphi_{i,j+d} = \varphi_{i,j+d_{max}}$), and all neurons are trained to output d=d_{max} (except those in columns 1 to d_{max}).

During testing, the pixels of I_R are copied to Φ. Then, each $n_{i,j}$ neuron output is computed, which is a disparity estimate for the $\varphi_{i,j}$ pixel of I_R. The disparity map is straightly given by the network output.

IV. Experimental Methodology

A. Datasets

To evaluate the performance of VG-RAM WNN on stereo matching, we used the Middlebury Stereo Datasets (http://vision.middlebury.edu/stereo/data/) [13, 14]. These datasets are composed of synthetic stereo image pairs with hand-labelled ground-truth disparities. In order of comparing our resultswith those submitted to the Middlebury Stereo Evaluation system (http://vision.middlebury.edu/stereo/eval/), we used the following four stereo image pairs of the Middlebury Stereo Datasets: "Tsukuba" and "Venus", from the 2001 datasets, and "Teddy" and "Cones", from the 2003 datasets.

B. Metric

To examine the performance of VG-RAMWNN on stereo matching, we used the *percentage of bad matching pixels*[20], because it is the metric adopted by the Middlebury Stereo Evaluation system. The percentage of bad matching pixels is given by:

$$B = \frac{1}{N} \sum_{(x,y)} \left(\left| d_C(x,y) - d_T(x,y) \right| > \delta_d \right)$$

whereδ_d is a disparity error tolerance, d_C is the computed disparity map, d_T is the ground truth disparity map and N is the number of pixels. In our experiments, we used δ_d= 1, since this concides with the Middlebury Stereo Evaluation system.

We evaluated the*average percentage of bad pixels*, that is given by the average of the percentage of bad matching pixels for the whole disparity map (*all*), non-occlusion (*nonocc*) regions and discontinuity regions (*disc*), i.e., the object boundaries.

V. EXPERIMENTAL RESULTS

In this section, we present the experiments employed to evaluate experimentally the performance of VG-RAM WNN on stereo matching.

The VG-RAM WNN architecture for stereo matching has two parameters: the number of synapses per neuron, |W|, and σ (see Section III). (Note that the number of neurons and the size of the network input must be equal to the size of the input image.) To tune the parameters of the VG-RAM WNN architecture, we trained it with the right images of the stereo image pairs and tested it with the left images, while varying the number of synapses per neuron and the σ value.We tested networks with number of synapses per neuron equal to 16, 32, 64, 128, 256, 512 and 1024, and σ equal to 1, 2, 4, 6, 8 and 10.

Figure 3 to Figure 6 present the results of the experiments we carried out to tune the parameters of the VG-RAM WNN architecture for stereo matching using the “Tsukuba”, "Venus", "Teddy" and "Cones" stereo image pairs, respectively. As Figure 3 to Figure 6 show, the performance (in terms of the average percentage of bad matching pixels) of the VG-RAM WNN architecture improves with σ; however, as σ increases, the performance stabilizes (for “Tsukuba” and "Venus") or even deteriorates (for "Teddy" and "Cones)". In the one hand, for smaller σ values, the synaptic interconnection distribution of neurons is concentrated on a smaller region of the network input, which limits the amount of available information for neurons. In the other hand, for larger σ values, the synaptic distribution is dispersed across a larger region of the network input and neurons may lose discriminative regions. The best and simplest (smallest σ) configuration is reached around σ=4, for all stereo image pairs examined.

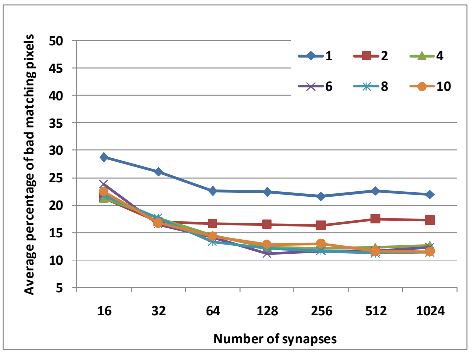

Figure 3: Performance tuning for the “Tsukuba” stereo image pair.

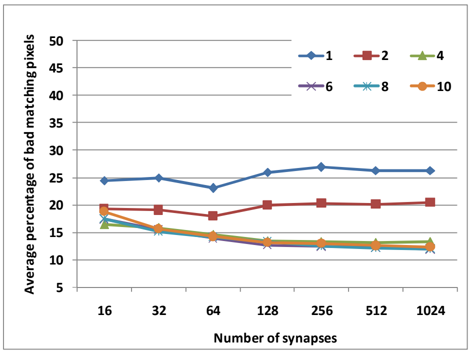

Figure 4: Performance tuning for the “Venus” stereo image pair.

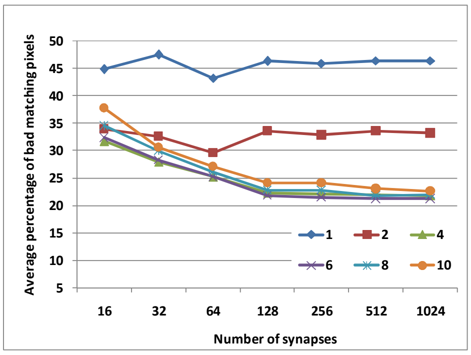

Figure 5: Performance tuning for the “Teddy” stereo image pair.

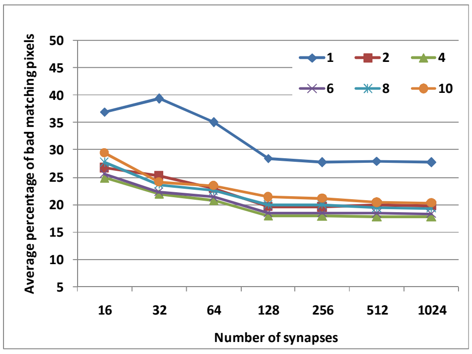

Figure 6: Performance tuning for the “Cones” stereo image pair.

As Figure 3 to Figure 6 also show, for σ=4, the performance of the VG-RAM WNN architecture improves with the number of synapses per neuron; however, as the number of synapses increases, the gains in performance decrease towards the maximum performance. For a small number of synapses, the number of elements of the network input monitored by neurons is small, which limits the amount of available information for neurons. The best and simplest (smallest number of synapses) configuration has around 256 synapses, for σ=4 and for all stereo image pairs examined.

The best and simplest VG-RAM WNN architecture configuration (σ=4 and 256 synapses) presented a performance (in terms of the average percentage of bad matching pixels) of approximately 12%, 13%, 21% and 14.5% for the “Tsukuba”, "Venus", "Teddy" and "Cones" stereo image pairs, respectively.

Figure 7 to Figure 10 show the right image and the disparity map generated by the best VG-RAM WNN architecture configuration for the “Tsukuba”, "Venus", "Teddy" and "Cones" stereo image pairs, respectively. These disparity maps were submitted to the Middlebury Stereo Evaluation system on April 12th 2012, being ranked at the 114th position. Figure 11 shows the result of this submission, that was partitioned into Figure 11(a) and Figure 11(b) for better readability. In Figure 11, column 1 presents the stereo matching algorithm submitted to the Middlebury Stereo Evaluation system (our system was listed at the third line); column 2 presents the average rank over the twelve succeeding columns, by which the table is sorted; columns 3-5 present the percentage of bad matching pixels along with the rank positionfor the nonocc, all and disc regions, respectively, for the “Tsukuba” image; columns 6-8, 9-11, 12-14 present analogous information for the "Venus", "Teddy" and "Cones" images, respectively; finally, the column 15 presents the average over all the twelve preceding columns.

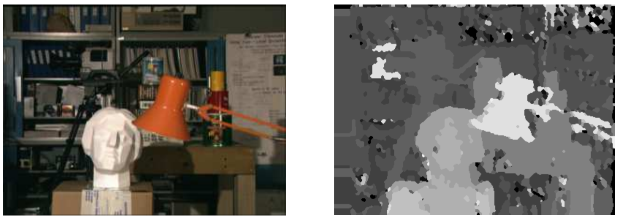

Figure 7: Right image of the stereo pair and the disparity map generated by the best VG-RAM WNN architecture configuration using the “Tsukuba” stereo image pair.

Figure 8: Right image of the stereo pair and the disparity map generated by the best VG-RAM WNN architecture configuration using the “Venus” stereo image pair.

Figure 9: Right image of the stereo pair and the disparity map generated by the best VG-RAM WNN architecture configuration using the “Teddy” stereo image pair.

Figure 10: Right image of the stereo pair and the disparity map generated by the best VG-RAM WNN architecture configuration using the “Cones” stereo image pair.

Algorithm	Avg. Rank	Tsukuba ground truth nonocc	Tsukuba all	Tsukuba disc	Venus ground truth nonocc	Venus all	Venus disc
2DPOC [110]	108.8	2.88 85	4.80 96	10.5 80	6.55 115	7.82 116	17.4 104
Bipartite [78]	110.3	2.54 79	4.41 90	13.6 94	6.62 116	7.46 115	18.6 107
YOUR METHOD	113.0	9.18 127	10.7 127	20.3 118	7.91 119	8.88 118	22.2 113
PhaseBased [31]	113.3	4.26 107	6.53 113	15.4 103	6.71 117	8.16 117	26.4 119
BioDEM [117]	114.6	6.57 124	8.43 123	28.1 126	3.61 110	4.80 110	33.7 122
RegionalSup [38]	114.8	3.99 103	6.05 108	14.2 95	8.14 121	9.68 122	36.8 126
IMCT [62]	114.8	4.54 108	5.90 107	19.8 117	3.16 107	3.83 106	23.2 115
SSD+MF [1a]	115.3	5.23 118	7.07 115	24.1 124	3.74 111	5.16 112	11.9 89
SO [1c]	117.2	5.08 115	7.22 119	12.2 88	9.44 124	10.9 124	21.9 112
GANCM [128]	118.8	4.82 112	6.81 114	16.6 107	8.07 120	9.54 120	36.0 125
MI-nonpara [85]	119.3	5.59 119	7.54 121	18.8 114	7.50 118	8.99 119	35.0 123
PhaseDiff [23]	120.5	4.89 114	7.11 117	16.3 105	8.34 123	9.76 123	26.0 118
STICA [16]	120.8	7.70 125	9.63 126	27.8 125	8.19 122	9.58 121	40.3 127
Rank+ASW [84]	120.8	6.51 123	8.43 123	19.7 116	10.5 126	12.0 126	32.7 121
CDM+AdaptWgt [75]	121.0	5.98 122	7.84 122	22.2 122	14.5 127	15.4 127	35.9 124
Infection [10]	122.3	7.95 126	9.54 125	28.9 127	4.41 113	5.53 113	31.7 120

(a)

Teddy ground truth nonocc	Teddy all	Teddy disc	Cones ground truth nonocc	Cones all	Cones disc	Average percent of bad pixels (explanation)
14.4 112	22.1 111	27.9 114	15.2 126	22.7 124	24.5 122	14.7
16.9 120	24.1 118	30.2 117	15.1 125	21.8 123	23.0 119	15.4
13.6 110	22.4 112	28.6 115	6.12 95	14.7 101	15.8 101	15.0
14.5 113	23.1 113	25.5 106	10.8 118	20.5 120	21.2 114	15.3
13.2 109	21.3 109	34.5 121	6.84 104	16.0 107	19.8 110	16.4
18.3 124	26.7 124	32.1 118	9.16 110	19.3 116	19.9 111	17.0
18.0 123	23.1 114	35.3 122	12.7 120	18.5 114	27.9 125	16.3
16.5 118	24.8 119	32.9 120	10.6 117	19.8 117	26.3 123	15.7
19.9 125	28.2 127	26.3 109	13.0 122	22.8 125	22.3 116	16.6
16.6 119	25.2 122	38.7 126	10.9 119	20.7 121	24.3 121	18.2
17.4 121	25.7 123	36.9 123	10.2 115	19.9 118	22.6 118	18.0
20.0 126	28.0 126	29.0 116	19.8 127	28.5 127	27.5 124	18.8
15.8 115	23.2 115	37.7 124	9.80 112	17.8 111	28.7 126	19.7
15.7 114	24.1 117	32.8 119	14.1 123	23.1 126	21.7 115	18.4
20.8 127	27.3 125	38.3 125	8.90 109	17.2 110	20.0 112	19.5
17.7 122	25.1 121	44.4 127	14.3 124	21.3 122	38.0 127	20.7

(b)

Figure 11: Result of the submission to the Middlebury Stereo Evaluation system of the disparity maps generated by our VG-RAM WNN architecture for stereo matching.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we presented an experimental evaluation of the performance of Virtual Generalizing Random Access Memory Weightless Neural Networks (VG-RAM WNN) on binocular dense stereo matching. We examined its performance with the Middlebury Stereo Datasets. Our experimental results showed that our VG-RAM WNN architecture for stereo matching was able to rank at the 114th position in the Middlebury Stereo Evaluation system. This result is promising, because our approach has not tackled occlusions and discontinuities in the stereo image pairs examined. Also, the difference in performance among approaches ranked in distinct positions is very small.

A direction for future work is to perform experiments with the KITTI Vision Benchmark Suite [21], which is suitable for analysis of disparity maps from real world scenes. Other direction for future research is to examine mechanisms for dealing with occlusions and discontinuities in stereo image pairs.

VII. ACKNOWLEDGMENTS

Wewouldliketothank Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq-Brasil (grants 552630/2011-0, 308096/2010-0, and 314485/2009-0) and Fundação de Amparo à Pesquisa do Espírito Santo-FAPES-Brasil (grant 48511579/2009) for theirsupporttothisresearchwork.

VIII. REFERENCES

[1] S. Thrun, W. Burgard, D. Fox, “Probabilistic Robotics”, MIT Press, 2005.

[2] K. Zhang, J. Lu, G. Lafruit, “Cross-Based Local Stereo Matching Using Orthogonal Integral Images”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, no. 7, pp. 1073-1079, 2009.

[3] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, X. Zhang, “On Building an Accurate Stereo Matching System on Graphics Hardware”, Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference, pp.467-474, 2011.

[4] Q. Yang, L. Wang, N. Ahuja, “A Constant-Space Belief Propagation Algorithm for Stereo Matching”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1458-1465, 2010.

[5] J. Sun, N. Zheng, H.Y. Shum, “Stereo Matching Using Belief Propagation”, Pattern Analysis and Machine Intelligence, vol. 25, no. 7, pp. 787–800, 2003.

[6] M. Vanetti, I. Gallo, E. Binaghi, “Dense Two-Frame Stereo Correspondence by Self-organizing Neural Network”, Proceedings of the 15th International Conference on Image Analysis and Processing (ICIAP '09), pp. 1035-1042, 2009.

[7] J. A. Martins, J. M. F. Rodrigues, J. M. H. du Buf, “Disparity Energy Model Using a Trained Neuronal Population”, Proceedings of the 2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 287-292, 2011.

[8] A. F. De Souza, C. Badue, F. T. Pedroni, E. Oliveira, S. S. Dias, H. Oilveira, S. F. Souza, “Face Recognition with VG-RAM Weightless Neural Networks”, Proceedings of the 18th International Conference on Artificial Neural Networks (ICANN 2008), pp.951-960, 2008.

[9] C. Badue, F. T. Pedroni, A. F. De Souza, “Multi-label Text Categorization Using VG-RAM Weightless Neural Networks”, Proceedings of the 10th Brazilian Symposium on Neural Networks, pp.105-110, 2008.

[10] H. P. Filho, A. F. De Souza, “VG-RAM WNN Approach to Monocular Depth Perception”, Proceedings of the 17th International Conference on Neural Information Processing: Models and Applications (ICONIP'10), pp. 509-516, 2010.

[11] I. P. Howard, B. J. Rogers, “Binocular Vision and Stereopsis”, New York: Oxford University Press, 1995.

[12] I. Aleksander, “From WISARD to MAGNUS: A Family of Weightless Virtual Neural Machines”, RAM-Based Neural Networks, World Scientific, pp. 18-30, 1998.

[13] D. Scharstein, R. Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms”, International Journal of Computer Vision, vol. 47, no. 1/2/3, pp. 7-42, 2002.

[14] D. Scharstein, R. Szeliski, “High-Accuracy Stereo Depth Maps Using Structured Light”, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), pp. 195-202, 2003.

[15] I. Aleksander, “Self-Adaptive Universal Logic Circuits”, IEEE Electronic Letters, vol. 2, no. 8, pp. 231-232, 1966.

[16] T.B. Ludermir, A.C.P.L.F. Carvalho, A.P. Braga, M.D. Souto, “Weightless Neural Models: A Review of Current and Past Works”, Neural Computing Surveys, vol. 2, pp. 41-61, 1999.

[17] G. R. Bradski, A. Kaehler, “Learning OpenCV: Computer Vision with the OpenCV Library”, Sebastopol, CA: O'Reilly Media, 2008.

[18] E. R. Kandel, J. H Schwartz, T. M. Jessell, “Principles of Neural Science”, Prentice-Hall International Inc, 2000.

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

[20] D. Scharstei, R. Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms”, International Journal of Computer Vision, vol. 47, no. 1-3,pp. 7-42, 2002.

[21] A. Geiger, P. Lenz, R. Urtasun, “Are we ready for Autonomous Driving?”, Proceedings of the Computer Vision and Pattern Recognition (CVPR), pp. 3354 – 3361,2012.