Overcoming the Challenge of Texture Classification Using Neuromorphic Hardware

Anne Menendez, Guy Paillet, Stephane Gervais-Ducouret
General Vision, 94952, CA, USA
anne@general-vision.com, guy@general-vision.com, sgd@general-vision.com
Paper written for EPoSS Annual Forum, October 19th, 2017
European Technology Platform on Smart Systems Integration

Abstract – **Image analytics can achieve great results in term of accuracy, but it is still far from becoming ubiquitous due to its requirements in terms of hardware and implementation costs. Current solutions are based on complex algorithms running on high performance Von Neumann hardware architectures, and have serious limitations in terms of power consumption, portability, latency, cost and deployment. A very different method is proposed using a parallel neural network chip. This can be described as a trainable, non-linear classifier enabling the development of imaging systems with a small foot print, low-power consumption and portability. Because this neural network is natively parallel and scalable, its latency to recognize a pattern is independent of the size of the knowledge stored in the neurons. This capability is a paradigm shift for the development of imaging systems where learning by example can lead to the generation of massive amounts of models to be examined. In this paper we elaborate this subject using the example of texture classification which is known to be fuzzy and difficult to model, and particularly difficult in detecting surface defects.**

***Keywords* - Neuromorphic; Image recognition; Image analytics; Texture classification; Pattern recognition; Cognitive Computing;**

I. INTRODUCTION

The demand for efficient image recognition in our day to day activities is everywhere, from healthcare monitoring, video surveillance, security, gaming, automotive ADAS (Advanced Driver Assistance Systems) and industrial automation. During the past 30 years [1], image processing has greatly evolved thanks to the increased performance of processors that execute compute intensive algorithms such as Wiener filter, Hull algorithm, Hough and Wavelet Transforms, Scale Invariant Feature Transform (SIFT) and other structural and feature space modeling methods [2].

These methods have enabled the deployment of advanced and high speed imaging applications in markets which can afford heavy duty and expensive systems, such as industrial inspection or intelligent transportation systems. However, some constraints of existing methods are preventing their broader adoption in markets that are power and cost sensitive, for example, in the consumer, automotive and mobile markets. A major limitation is the cost of these systems, which is partly related to the hardware and software development, but also to the accessories and logistics necessary to control the environment and ensure a successful outcome of the algorithm. This can imply expensive lighting or specific conveying of objects to separate them or control of the object orientation, etc.

One difficult subject in image recognition is texture classification, which is actually a necessary step leading to image segmentation and context understanding, and is used in applications such as medical imaging, aerial and satellite imaging, as well as food and raw produce inspection. A solution with small foot print and low power could be adopted in numerous appliances and small devices, for instance, dermatological scanners, monitoring of crops by drones, handheld electronic field guides, fauna and flora recognition, etc.

In this paper, we are proposing a new method for texture classification and anomaly detection based on a neuromorphic chip called NeuroMem. This chip can be labeled as a neuromorphic memory capable of learning and recalling a pattern, but it is also a neural network since it has a parallel architecture where memory and recognition logic are tightly coupled together, and the recognition latency is independent of the number of reference patterns to be compared. Neural networks were invented 50 years ago [3] and software versions have been used with success in many imaging applications from face recognition, fish inspection to OCR and more [4], [5], [6]. The availability of a fully parallel neuromorphic chip [7] enables the use of a neural network and particularly, the powerful non-linear Radial Basis Function classifier offering a small foot print, low-power consumption and real-time learning capability. Also, deploying the image recognition engine does not require any complex software nor high-end computing power, but rather a graphical interface to annotate and teach reference models [8].

In this paper, we also describe the various benefits of a parallel hardware neural network for the classification of texture, and specifically the example of surface anomaly detection.

II. CHALLENGES OF TEXTURE CLASSIFICATION

The representation of a texture can have many different models characterizing its color, granularity, and dominant edge orientation. The descriptor of a small patch of texture can be its raw pixel values or a signature extracted from these pixel values which gives insights about the texture. Depending on the type of surface to classify (a landscape, a material, etc) and its variability, hundreds to several thousands of examples may be necessary to model its similarities and scarcities properly. The decision space can be quite complex with disjoint, embedded and overlapping categories as shown on figure 1.

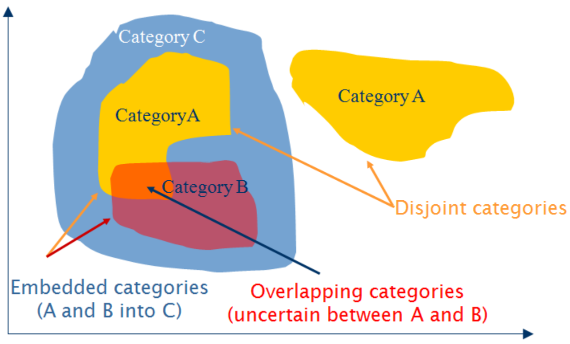

Figure 1: Example of a complex 2D decision space

The examples retained as references constitute the knowledge base. In the NeuroMem neural network, each neuron can hold in its memory the image of a patch of texture and a label assigned to this texture at the time of its learning (an example of a label in the case of land surveying: grass, forest, sand). When a new patch is broadcasted to all the neurons in parallel, the ones recognizing it as similar to their model in memory will fire, and a Winner-Takes-All race will occur to deliver the best result in a constant amount of time (see fig. 1).

A. *Limitations of Standard Methods*

The aim of image recognition is to extract significant information within a flow of pixel data and extract multiple signatures to verify their conformity with the signatures of reference models. Numerous books and papers discuss the data mining challenges and proposed improvements to cope with the massive amount of images generated today [9], [10] [11]. Current imaging methods clearly require high-performance computation platforms to handle large matrix calculations, filtering, morphological transformations, feature extractions, blob measurements, etc. Also, with the Von Neumann architecture, pattern recognition time drastically increases with the number of models to compare since memory and computing are compartmentalized which results in bottle necks such as memory bandwidth, fetch and decode latency and hence, an inadequate architecture. Even with new architectures that combine GPU (Graphic Processing Unit) and multi-core processors [12] , the speed of pattern matching is not satisfactory in many cases. The cost of the equipment and power consumption level are not practical, prohibiting portability and battery-operation if needed.

B. *Ideal System Characteristics*

Looking at the targeted applications listed above for industrial, medical and automotive markets, the ideal vision system, to become pervasive, needs to reach high response time speed while consuming low power with a small foot print, fast implementation and lower cost. Hence, the following requirements at the solution level can be derived and listed as follows:

- Extremely small time-delay and a predictable delay which can allow near real-time response.
- Portability. Small foot print and low-power consumption are paramount for portability, flexibility and wide adoption at device levels. Power consumption must be small enough to allow battery operation and embedded image recognition in small spaces such as a car or wearable equipment.
- Trainability: The system should be able to learn new patterns and be adaptive
- Clonability: Knowledge must be easily reproducible and cloned.
- Ease of use: Non-requirements of computer expertise and removal of the software complexity burden.

A neural network approach can address all of these requirements.

III. Texture Classification with Neural Network

A. *Suitability of the NeuroMem network*

Given that the neurons of the NeuroMem chip each have a memory of 256 bytes, they can fit a patch of 16x16 pixels. Also, the use of larger patches of pixels is possible if the signature extracted from these patches fit on 256 bytes. A single chip has 1024 neurons and can therefore hold up to 1024 different models. Considering that the average resolution of images has reached the 2Kx2K pixels, the partitioning of the image into patches of 16x16 pixels amounts to 16,384 adjacent patches and each patch must be compared to the models held in the neurons. If for example the knowledge built into the neurons amounts to 800 models, the processing of the image will require over 13 million comparisons. Sustaining a frame rate of 30 to 60 per second requires a processor delivering Giga operations per second. In the case of the parallel neural network, the number of comparisons will be the same except that the time to find the best match is 10 microseconds, irrespective of whether the knowledge is composed of 800 or 800,000 models.

The natively parallel architecture of the NeuroMem chip also implies that a chain of multiple chips can be easily assembled by connecting them together, thus increasing the number of neurons available to store models by 1024 per chip. The NeuroMem chip is the successor of the IBM ZISC® chip (Zero Instruction Set Computer) invented and patented in 1993. Regardless of the number of these chips daisy-chained together, they will make a large bank of identical neurons which can all learn and interact with each other to produce a global response with regards to the recognition of an incoming pattern or, if applicable, its identification as a new pattern. The chain of chips operates at 16 MHz and the recognition time is 10 µsec per pattern of 256 bytes.

The **latency being independent of knowledge capacity** is a valuable attribute since recognition speed is not impacted by the number of models necessary to represent single or multiple textures. Consequently, it is possible to use 16x16 raw pixel values to model textures without any need for either data conditioning or regards to the amount. The learning process becomes very easy and fast as opposed to the extraction of sophisticated features such as wavelet and FFTs.

The neurons can have two different behaviors both important for texture classification: *K*-Nearest Neighbor algorithm (*k*-NN) and Radial Basis Function (RBF) [13].

The **K-Nearest Neighbor** algorithm (k-NN) classifies patterns by reporting their K closest neighbors stored in the neuron memory. It can be useful to segment and clusterize an image, but it is useless in the case of anomaly detection since it always produces an output (the closest) and does not have any notion of "unknown".

The **Radial Basis Function** has a different activation function and can classify objects with the notion of (1) unknown or novelty, (2) positive identification, or (3) identification with uncertainties where multiple categories are identified. The acknowledgment of uncertainties is very important since it allows triggering decision rules or hypotheses, making the classification more robust. This can be especially powerful for applications where the cost of the mistake is high such as in medical imaging.
The architecture of a single neuron is represented figure 2 with the different parameters.

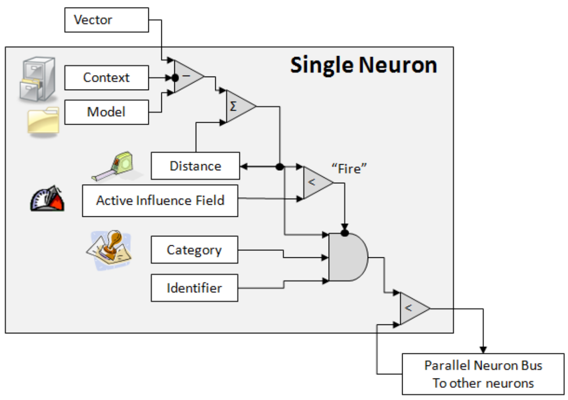

Figure 2: Architecture of a single neuron

Regardless of the selected classifier (BRF or KNN), the neurons of the NeuroMem chip have the following behavior during the broadcast of a pattern:

- Firstly, all the neurons update their distance register simultaneously as they receive the next component of the input pattern broadcasted on the parallel neuron bus. Upon receipt of the last component, all neurons have their distance calculated and "fire" if this value is smaller than their Active Influence Field. For example, if an input vector has 200 bytes and is broadcasted to a network of 10,240 neurons, their 10,240 distance values are calculated after the 200th clock cycle and the network raises its "Identified" line if at least one neuron is firing. The smaller the distance value, the closest the match between a model and the input pattern [14].
- Secondly, the firing neurons are autonomously capable of ordering themselves per increasing value of distance if a host processor sends successive requests to read their response (i.e. distance, category or identifier). At each read command, only the next neuron with the smallest distance responds. This unique self-discipline pertains to the parallel architecture of the NeuroMem chip and its patented "search and sort process" [15].

B. Texture modeling with RBF classifier

Neurons are taught by example. If the host processor broadcasts a pattern followed by a "Learn" command accompanied by the category to learn, the next available neuron in the chain stores the pattern and category if, and only if, it is not recognized by any committed neuron. For example, in wood sorting, categories can be "loose grain", "fibrous" and "knotted". In the case of surface defect classification, categories can be "scratch", "hole", "spot", etc. Finally, in the simple case of anomaly detection, categories can be as simple as "good" and "bad".

The teaching interface can be designed to be supervised by an operator, or semi-supervised through the execution of a high level function in charge of assigning a category value to newly learned patches of texture. The neural network can be set to learn models in a more or less conservative manner at the neuron level. In the case of a conservative option, any slight variation from the good models will be considered as a defect. The throughput of a production line might drop but the quality will be very high. On the contrary, if the neurons are taught under a liberal mode, which means that they are authorized to perform a larger degree of generalization, some anomalies might not be detected, but a customer might require that if the throughput of the production line is a priority of the moment or the season.

Building a rich and accurate knowledge base requires the learning of many images of texture. The number of committed neurons should increase following an exponential curve at first and then stabilize as enough examples have been learned.

Note that since the number of learned models does not impact the response time of the network, the patches of texture can be learned at every pixel position in the annotated region. For example, if the region covers an entire image of 752x480, the later represents 47x30 or 1410 blocks of 16x16 pixels. The first 47x30 blocks are located at an offset (0,0) from the upper left corner of the image. Then 46x30 examples can be extracted with an offset ranging from 1 to 15 along the horizontal axis. This shift can be repeated 15 times along the vertical axis. The total number of samples to learn becomes greater than 300,000 per image, but the number of samples retained by the neurons should be much smaller since generalization is supposed to occur.

Moreover, teaching of additional textures can be done at any time. So a knowledge base can be fine tuned over multiple cycles of production. It becomes the intellectual property for the user after a while.

C. Texture classification with RBF classifier

The recognition of a pattern has been described in section A and its outcome can be one of the following three statuses: (1) Identified, (2) Uncertainty or (3) Unknown. Several neurons can recognize the same input pattern. The one with the smallest distance value holds the pattern which is the closest to the input vector. The neuron with the 2^{nd} smallest distance value holds the pattern which is the next closest to the input vector. It is possible that multiple neurons recognize the vector with the same smallest distance. If they have identical categories, it reinforces the confidence level of the recognition. If they do not have the same category, they point to a level of uncertainty and potential ambiguities between categories. This uncertainty can be further considered by reading the categories recognized by the next firing neurons and applying decision rules to come up with the most probable category. Rules have to be established on a “per application” basis depending on the cost of a mistake, the requirements for a minimum throughput, minimum false negative, etc.

D. NeuroStack, highly scalable platform

The NeuroStack board of General Vision has been designed to easily size a chain of NeuroMem chips by simply stacking boards on top of each other. Each individual board features a chain of four chips, or 4096 neurons, and a Field Programmable Gate Array (FPGA) which acts as the glue logic between the neurons, external sensors, actuators and communication ports. In the case of texture classification, the FPGA is in charge of extracting the patches of 16x16 pixels from the image, broadcasting them to the neurons and reading the response of the firing neurons to build a final decision and transmit it to a control system or PLC (Programmable Logic Controller) or actuator. The pixel data can be received in real-time from a sensor or pre-loaded to the board’s memory. If an application requires more neurons, additional boards can be stacked in which case their FPGA is just used to establish the chain between the NeuroMem chips. After stacking N boards, the recognition of a patch of 16x16 pixels still takes 10 microseconds against 4096 times N possible models and the power consumption of the stack is N times 2 watts.

IV. Application of Industrial Surface Defect Detection

Surface Defect Detection is a subset of texture classification where the category of a texture is irrelevant as long as it is of a “known” type. It is a common and difficult entry application in industries such as paper and pulp, lumber and glass. The material to inspect is highly textured and the system has to find small abnormalities within a texture which can be patterned or highly irregular. Some examples of patterned texture are represented on figure 3.

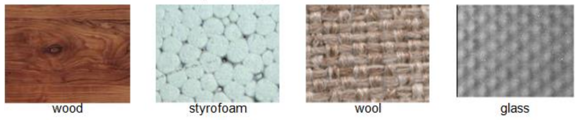

Figure 3: Example of patterned textures

For Surface Anomaly Detection, the recognition engine is programmed to monitor if the neurons do NOT recognize a patch of texture, meaning that it represents a novelty and possible anomaly. If such event occurs, the application can simply report the anomaly and its location, or it can store the pixels of the patch for later review by a human supervisor who can decide to add the patch as a good example, if applicable.

In the case of industrial web inspection, the sensors are aligned on top of the float or conveyor on which the material to inspect flows at a constant speed. Inspected material can be glass, wood, vinyl, pulp paper, etc. Current systems involve expensive cameras, fixture and lighting mounted over the moving material and connected to one or multiple high performance computers in charge of processing the images at high speed but at high cost and power consumption. One significant portion of the cost is the programming of the multiple processing stations. With a NeuroMem-powered sensor, the approach is different and simpler since the inspection is made by an array of identical smart sensors covering the width of the material to inspect and capable of making decisions autonomously.

A. CogniSight in-line sensors

The CogniSight Sensor of General Vision has been developed to easily construct a chain of identical sensors, which each have their own NeuroMem network, but loaded with the same knowledge base. The sensor is a monochrome CMOS sensor with a resolution of 752x480 pixels running at 60 frames per second, weighs less than 120g and its power consumption is about 0.5 Watt. It can detect anomalies as small as 2x2 pixels in less than 75 msec. The sensors are daisy-chained over an RS485 line so a host controller can ping them separately or all at once. Figure 4 shows the CogniSight sensor array and Figure 5, the overall system.

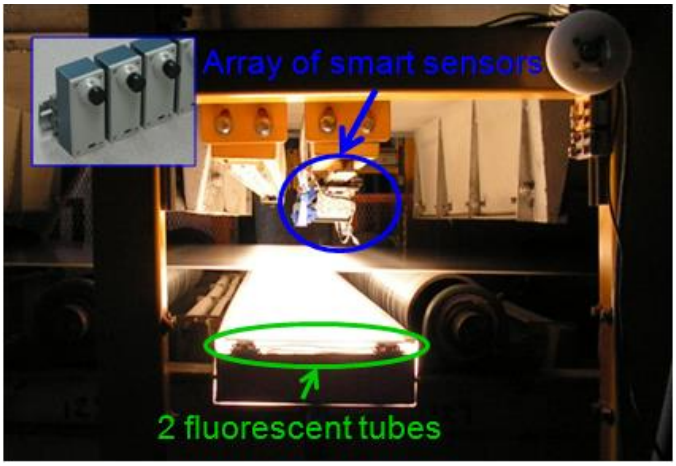

Figure 4: Array of sensors mounted above the glass float and standard fluorescent lighting mounted underneath

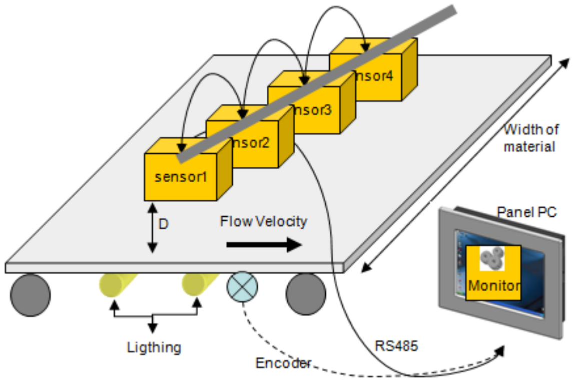

Figure 5: System diagram

B. Off-line primitive training, on-line tuning

The training of what is an acceptable surface is performed using images collected on the real production line and with the sensors mounted at their final location and height from the flowing material. The training is performed off-line with a simple user interface in which images of good and bad texture are annotated. Typically, the operator has to review the collected images and tag them as showing an acceptable surface, or conversely, a non-acceptable surface, and do this without specifically pointing at the anomalies. The program uses the tagged images and the model generator of the neurons to build a decision space, or knowledge base, which properly identifies the anomalies in the non-acceptable samples. A knowledge can be qualified as valid once new examples of texture are recognized as “Identified”, and invalid when defective patches are not recognized at all (i.e. status “Unknown”). Different knowledge bases can be built and saved per type of material, type of client, etc. Depending on the installation and type of material to inspect, it might be necessary to teach more examples to the sensors at each end of the line since they have to deal with the edge patterns of the material. In the case of patterned glass, a learning phase may be executed over several days of production to generate a knowledge base of 800 neurons, so as to properly detect scratches and bubbles.

C. Recognition

At each ping from the host controller, all sensors transmit a minimum of one byte of acknowledgement. The sensors having detected an anomaly in their field of view transmit a few additional bytes (sensor ID, X and Y position of the defect). This takes 4 microseconds over a serial line and the data can be sent to a marker, cutter or other actuator down the production line. Depending on the velocity of the material flow, the speed performance of the system may allow for multiple detections of an anomaly in the sensors’ field of view thus reinforcing the robustness of the detection.

The plurality of identical sensors makes them easy to install and maintain. Hence, if the ping of the sensors reveals that a sensor is not responding, it is easy to slide out the DIN rail, replace the defective sensor and pull back the rail over the material.

Since the knowledge built by the neurons can be saved to a knowledge file, the system can be configured with different settings addressing different quality levels or target customers.

V. Perspectives

Texture classification and industrial surface defect detection systems with NeuroMem-based hardware can not only improve the effectiveness of existing vision systems but also reach new levels of performance with learning capability, modularity, higher responsiveness, lower cost and power consumption. This distributed and scalable hardware solution is also promising for other applications in video analytics, machine vision and smart miniature sensors.

Since the NeuroMem chips only require relatively modest semiconductor process technology that is several nodes less than the highest performance conventional microprocessors and memories, new successive generations are expected to have a much greater density of neurons and consume less power. This will enable even more sophisticated applications for surface defect detection and vision systems.

VI. Bibliography

[1] S. J., Image analysis and mathematical morphology, London New York: Academic Press, 1982.

[2] D. Luo, "Pattern Recognition and Image Processing," Woodhead Publishing Limited,, 2011, pp. 1-5.

[3] DARPA, "DARPA Neural Network Study," M.I.T. Lincoln Laboratory, Lexington, MA, 1988.

[4] I. Marosi, "Industrial OCR approaches: architecture, algorithms, and adaptation techniques," *Proceedings of SPIE* , vol. 6500, no. Document Recognition and Retrieval XIV, 2007.

[5] A. M. G. Paillet, "Fish Inspection System Using a Parallel Neural Network Chip and the Image Knowledge Builder Application," *Journal of AAAI,* 01 01 2008.

[6] C.-Y. Chang and S.-Y. F. H.-J. Wang, "Texture image classification using modular radial basis function neural networks," *Journal of Electronic Imaging 01/2010;,* vol. 19, no. 1, 2010.

[7] A. Menendez, "NeuroMem CM1K chip," General Vision, 13 May 2014. [Online]. Available: http://general-vision.com/cm1k/.

[8] A. Menendez, "Image recognition using neural networks," *RTC Magazine,* no. July 2001, 2001.

[9] L. Yaping, Z. Jinfang, X. Fanjiang and S. X., "The Recognition and Enhancement of Traffic Sign for the Computer-Generated Image," *Fourth International Conference on Digital Home (ICDH),* no. 23-25 Nov. 2012, pp. pp.405,410, 2012.

[10] F. S., W. L. and J. F., "Application of video image recognition technology in substation equipments monitoring," *International Conference on Electrical and Control Engineering (ICECE),* no. 16-18 Sept. 2011, pp. pp.4374,4377, 2011.

[11] S. Haddadi, C. Fernandez and F. Abdelnour, "Textural features and neural network for image classification," in *SPIE, Applications of Artificial Neural Networks in Image Processing*, San Jose, CA, 1996.

[12] S. Gervais-Ducouret, "M2M and Embedded Processing: Advantages of Multi-core & Asymmetric Processing," in *Proceeding of Embedded World Conference 2014*, Nuremberg, Germany, 2014.

[13] G. D. T. a. P. T. K. Madani, "Image Processing Using RBF like Neural Networks: A ZISC-036 Based Fully Parallel Implementation Solving Real World and Real Complexity Industrial Problems," *Applied Intelligence 18, 2* , pp. 195-213, 18 March 2003.

[14] A. M. G. Paillet, "Improved neuron circuit architecture". USA Patent US5717832, 10 02 1998.

[15] J.-Y. Boulet, P. Tannhof and G. Paillet, "Circuit for searching/sorting data in neural networks". USA Patent US5740326, 14 April 1998.