Neurocomputing 151 (2015) 11–33

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A bio-inspired multisensory stochastic integration algorithm

Alex Porras*, Rodolfo R. Llinas

Department of Neuroscience and Physiology, NYU School of Medicine, New York, NY 10016, USA

ARTICLE INFO

Article history:
Received 24 April 2014
Accepted 23 June 2014
Available online 6 November 2014

Keywords:
Bioinspired
Multisensory
Stochastic
SPGD
Intensionality vector
Motor output

ABSTRACT

The present paper describes a new stochastic multisensory integration system capable of combining a number of co-registered inputs, integrating different aspects of the external world, into a common premotor coordinate metric.

In the present solution, the model uses a Stochastic Gradient Descent (SGD) algorithm to blend different sensory inputs into a single premotor intensionality vector. This is done isochronally, as the convergence time is independent of the number and type of parallel sensory inputs. This intensionality vector, generated based on "the sum over histories" [1], makes this implementation ideal to govern non-continuous control systems. The rapid convergence of the SGD [2–7] is also used to compare with its biological equivalent in vertebrates -the superior tectum- to evaluate limits of convergence, precision and variability. The overall findings indicate that mathematical modeling is effective in addressing multisensory transformations resembling biological systems.

1. Introduction

Multisensory integration has only recently started to move beyond the exclusivity of the cortical mantle and it now accepts that other structures play a more formative role than the telencephalic sensory areas [8–20]. Thus, the V1 cortical area in visuali impared patients, may be activated by auditory, Braille reading or other somatosensory clues [14–21]. In addition, intense activity with superadditive properties [10] can occur in the superior collicullus due to the multisensory activation [18,22]. These integration sites must be considered in addition to the association cortices widely known for multisensory convergence [23].

It would be unwise, therefore, given the complexity of the CNS to address these questions all at once. However, in vertebrate evolution the optic tectum (superior colliculus) is a multisensory combination site [24–34] receiving audition, vision, somatosensory and vestibular inputs, where the inputs from each sensor modality are topographically organized. However, as these maps do not encode shape or size, but only spatial continuity and connectedness they are, from a mathematical perspective, topological rather than topographically organized [24,30]. The optic tectum, basically addresses multisensory representation in the context of posture and motricity, but most importantly, the multisensory stimuli is integrated to organize responses that exceed the sum of the individual stimuli [10,14] – this is called superadditivity.

With regard to defining "real time multiple sensory driven motor control systems," the issue has been hard to define [35] and most novel steps taken have been limited [36–41]. Common drawbacks relate to time response, output convergence, susceptibility to external noise, and most significantly, the proper "homogenization" of inputs from different sensors into a single intentionality vector.

This is essential in interfacing with the motor plant. In addition, from a dynamic perspective response, time is a non-linear function relating to the magnitude of the error control required [42–46]. Thus, sensory-motor convergence speed is directly related to the number of recurrent cycles (discrete method) required for response stabilization following sensory input arrival time [4–7,47,48].

1.1. Mathematical Model

Following the issue of supervised, unsupervised and gradient algorithms, a more precise definition of convergence speed in large and/or multidimensional data sets is necessary. A prerequisite is the use of an open or "unsupervised" algorithm during sensory input acquisition, to be normalized for generating a compound image. Following image generation via geometrically co-registered inputs, the resultant n-dimensional matrix becomes the supervised parameter on which the center of mass [49] and the local density distributions are analyzed.

The first part of the algorithm includes the normalization of the inputs. The basic initial assumption is that all the inputs are equivalent and that the multiplying factor k_i for each, is 1. This multiplying factor could be a variable filter if its variation can be

* Corresponding author.
E-mail address: alex.porras@nyumc.org (A. Porras).

http://dx.doi.org/10.1016/j.neucom.2014.06.080
0925-2312/

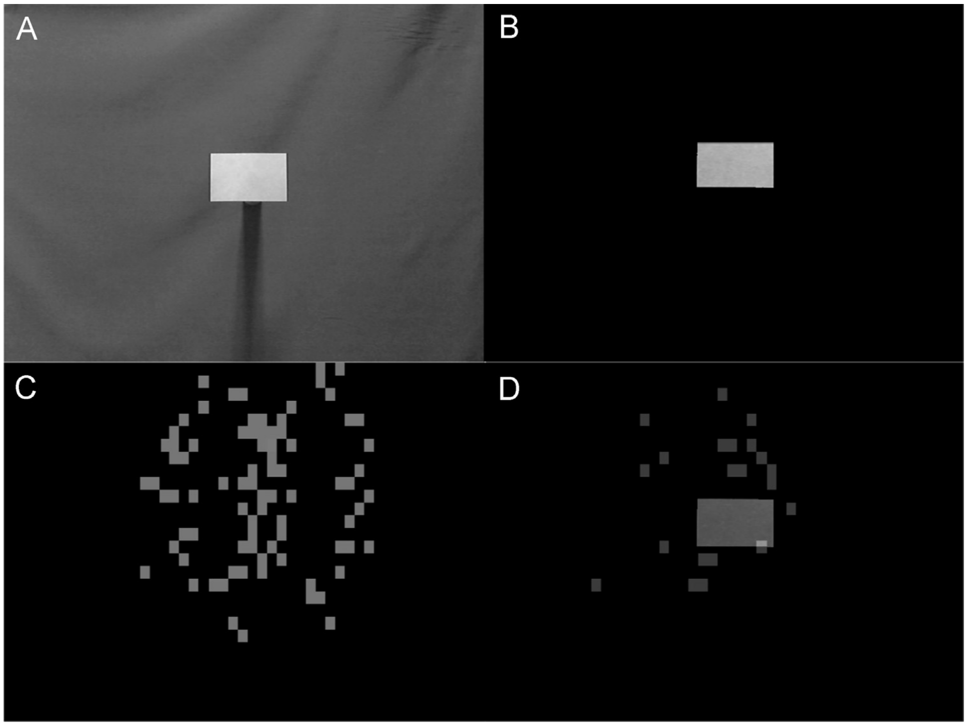

Fig. 1. A) Video camera image of a non-processed three-dimensional object. B) Filtered video image of the 3D object. C) Sonar image of the 3D object D) Combination of images of a 3D object seen by the camera and sonar simultaneously.

associated with external factors or with the algorithm's cost functions.

$$\bar{x}'(i,j) = k_n \cdot \bar{x}_n(i,j) \tag{1}$$

Such normalization procedure and the internal product for the factor k_i when applied to all the inputs allow their combination into one single array[13,50–54] the base image.

$$\bar{x}_m(i,j) = \frac{\bar{x}'_{n1}(i,j) + \bar{x}'_{n2}(i,j) + \dots \bar{x}'_{nm}(i,j)}{m} \tag{2}$$

As a first example of this procedure, the sonar and video input of a single rectangular object with a 3D projection (from the Materials and Methods section of this chapter) can be combined to observe the following results.

The results in Fig. 1 correspond to the co-registered image space of the two input sensors (camera and sonar). The mathematical analysis requires that, the visual dimensions of the figure width and height be used as the limit values of the input vectors x_{video}(width, height) and x_{sonar}(width, height), as measured in pixels.

The information received from the camera, after gray scale conversion and contrast increase filtering, is represented in each pixel as $x_{video}(i, j)$. See Fig. 1-B.

As with the camera output, the sonar output generates a different representation of the same object, as is the case in the biological counterpart with visual and auditory senses and a common stimulus. The array from the sonar has the same dimensions as the one from the camera, $x_{sonar}(i, j)$. See Fig. 1-C.

Thus far the unsupervised part of the algorithm generates the base image from the sensory inputs and will produce a normalized array of the same dimensions of the co-registered inputs of its summatory.

$$\bar{x}_m(i,j) = \sum_i^{weight} \sum_j^{height} \frac{k_{video}\bar{x}_{video}(i,j) + k_{sonar}\bar{x}_{sonar}(i,j)}{m} \tag{3}$$

Where,

$$m = k_1 + k_2 + \dots k_n \le n \tag{4}$$

This assumes that both inputs are equivalent during the analysis and the constants k_i have one (1) as a value.

From the previous summation equations, the resultant base image is shown in Fig. 1-D. At this point, the normalized image $\bar{X}_m(i,j)$ has the information necessary to start a simple search to locate the object or objects, which generate the densities in the image. As a control, the center of mass is obtained, employing as a weight the 8 bit gray scale of each pixel. Therefore,

$$mass(i,j) = Pixel_{grayscale}[\bar{x}_m(i,j)] \tag{5}$$

and in order to find the coordinates,

$$Center_{mass}(i) = \frac{\sum_{j=0}^{width}\sum_{i=0}^{height} j \cdot mass(i,j)}{\sum_{j=0}^{width}\sum_{i=0}^{height} mass(i,j)} \tag{6}$$

and for j,

$$Center_{mass}(j) = \frac{\sum_{i=0}^{width}\sum_{j=0}^{height} i \cdot mass(i,j)}{\sum_{i=0}^{width}\sum_{j=0}^{height} mass(i,j)} \tag{7}$$

The resultant base image with the center of mass located is represented on Fig. 2 (red circle).

Following the unsupervised center of mass acquisition phase, the supervised algorithm determines pixel number location to determine the maximum area density site.

The input array analyzed, the $\bar{X}_m(i,j)$, normalized sensor image, does not contain the expected solution to train a standard supervised algorithm. Instead, the system optimizes error determination based on the weight of all the gray scale pixels via a random sweep of the base image until converging into a maximum density area.

An arbitrary size rectangle P_{ij} of dimensions u and v is declared with origin coordinates given by a random generator algorithm,