Neurally Inspired Mechanisms for the Dynamic Visual Attention Map Generation Task

Maria T. López[1], Miguel A. Fernández[1], Antonio Fernández-Caballero[1],
and Ana E. Delgado[2]

[1] Departamento de Informática
E.P.S.A., Universidad de Castilla-La Mancha, 02071 – Albacete, Spain
{mlopez, miki, caballer}@info-ab.uclm.es
[2]Departamento de Inteligencia Artificial
Facultad de Ciencias and E.T.S.I. Informática, UNED, 28040 - Madrid, Spain
{adelgado}@dia.uned.es

Abstract. A model for dynamic visual attention is briefly introduced in this paper. A PSM (problem-solving method) for a generic "Dynamic Attention Map Generation" task to obtain a **Dynamic Attention Map** from a dynamic scene is proposed. Our approach enables tracking objects that keep attention in accordance with a set of characteristics defined by the observer. This paper mainly focuses on those subtasks of the model inspired in neuronal mechanisms, such as accumulative computation and lateral interaction. The subtasks which incorporate these biologically plausible capacities are called "Working Memory Generation" and "Thresholded Permanency Calculation".

1 Introduction

Visual attention models may be classified as space-based models – e.g., spotlight metaphor [12] and zoom-lens metaphor [1] - and object-based models – for instance, discrete objects [11, 14]. Our approach falls into the models based in objects, where visual attention always focuses on discrete objects or coherent groups of visual information. In order to select an object that captures our attention, we previously have to determine which image features should be combined to obtain the object. This process is carried out in two stages. Firstly, obtaining features to generate simple shapes, and, secondly, combining these simple shapes into more complex ones. Once the objects have been obtained, the next process –attention based on objects– consists in selecting one of the shapes generated.

Vecera [14] introduced a model to obtain objects separated from the background in static images by combing bottom-up (scene-based) and top-down (task-based) processes. The bottom-up process gets the borders to form the objects, whereas the top-down process uses known shapes stored in a database to be compared to the shapes previously obtained in the bottom-up process.

A first plausible neuronal bottom-up architecture was proposed by Koch and Ullman [9]. Their approach is also based in features integration [13]. In Itti, Koch and

Niebur [10] a visual attention system inspired in the neural architecture of the early visual system and in the architecture proposed by Koch and Ullman [9] was introduced. This system combines multi-scale image features in a unique saliency map. The most salient locations of the scene are selected from the highest activity of the saliency map using a winner-take-all (WTA) algorithm.

Another example based in the saliency map idea is the guided search (GS) model proposed by Wolfe [15]. This model incorporates two visual selection stages. The first one, the pre-attention stage, with a great spatial parallelism, performs the computation of simple visual features. The second stage is performed spatially in a sequential way, using more complex visual representations – including the combination of features - to perform the calculations. The value of the features obtained is a combination of bottom-up and top-down knowledge. Attention is directed to the location that shows the highest value in the **Dynamic Attention Map**. The guided search ends if the location contains the target looked for. If this is not the case, the search continues in the **Dynamic Attention Map** in a descending order, finishing when the target is found or when attention falls under a threshold.

In this paper we introduce a model of dynamic visual attention that combines bottom-up and top-down processes. Bottom-up is related to the first step of the architectures proposed, where the input image is segmented using dynamic criteria by means of neurally inspired accumulative computation [2-4] and lateral interaction [5-8]. The observer may indicate how to tune system parameters to define the attention focus using top-down processes. These processes are of a static nature, during the configuration of the features selection and the attention focus processing, or dynamic, when the observer modifies parameters to centre the focus on the interest elements.

2 Model of dynamic visual attention

Our approach defines a PSM for the generation of a **Dynamic Attention Map** on a dynamic scene, which is able to obtain the objects that will keep the observer's attention in accordance with a set of predefined characteristics. Figure 1 shows the result of applying our model to the "Dynamic Attention Map Generation" task, where attention has been paid on moving elements belonging to class "car".

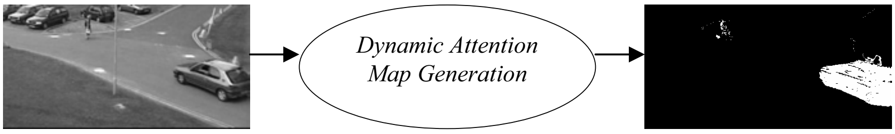

Fig. 1. Input and output of the "Dynamic Attention Map Generation" task

The different subtasks proposed to obtain the **Dynamic Attention Map** are depicted on figure 2, whilst figure 3 shows the inferential scheme of the model introduced. Next these subtasks are briefly described:

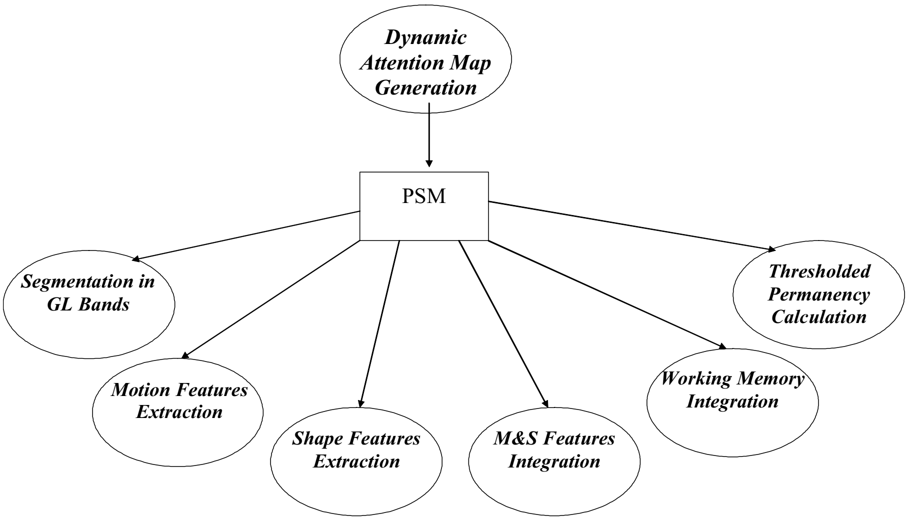
Dynamic Attention Map Generation
PSM
Segmentation in GL Bands
Motion Features Extraction
Shape Features Extraction
M&S Features Integration
Working Memory Integration
Thresholded Permanency Calculation

Fig 2. Decomposition of the "Dynamic Attention Map Generation" task into subtasks

a) Subtask "Segmentation in Grey Level Bands" is thought to segment the 256 grey level value input images into *n* (static role) grey level bands.
b) Subtask "Motion Features Extraction" calculates for each pixel a set of dynamic features, e.g. the presence or absence of motion, the speed, or the acceleration. The output of this subtask is formed by those pixels that deserve some dynamic features. Parameters M_i (static roles) indicate the range of values for motion features that pixels must possess to be marked as *activated* at the output of the subtask.
c) Now, subtask "Shape Features Extraction" calculates certain shape features such as the size, the width, the height, the relation width to height, the value size/(width*height), and so on. It obtains those elements that deserve a set of predefined features. Parameters S_i (static roles) indicate the range of values for shape features that elements must possess to be labelled as *activated* at the output of this subtask. The elements that do not have the features defined will be marked as *inhibited*. All the rest of pixels which do not belong to elements are labelled as *neutral*.
d) The idea of subtask "Motion & Shape (M&S) Features Integration" is to generate a unique **Motion and Shape (M&S) Interest Map** by integrating motion features and shape features in accordance with the criteria defined by the observer. These criteria correspond to C_i (static role).
e) Subtask "Working Memory Generation" segments the image in regions composed of connected pixels whose brightness pertains to a common interval. Each of these regions or silhouettes of a uniform grey level represents an element of the scene.

f) In subtask “Thresholded Permanency Calculation“ firstly the persistency of its input is accumulated. Then, a further threshold is performed. P_i are the static roles for this subtask.

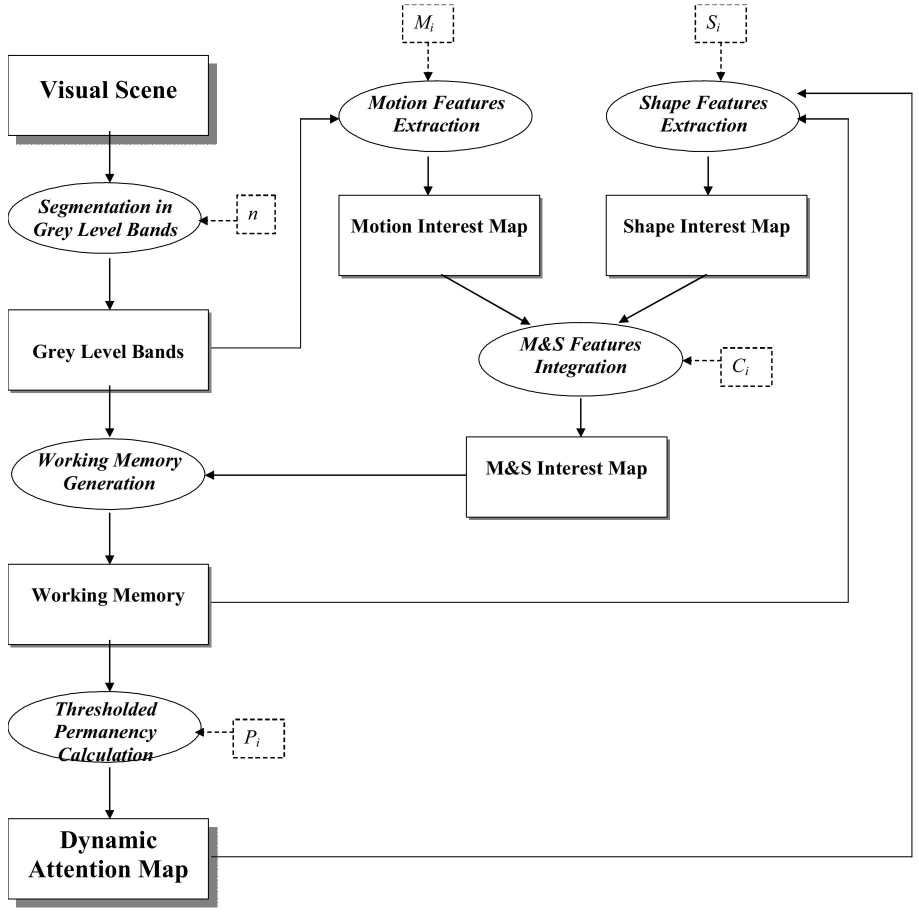

Fig 3. Inferential scheme of the model

From now on, this paper focuses on the generation of the working memory and the **Dynamic Attention Map** from the image segmented in grey level bands and the **Motion and Shape Interest Map**. As already mentioned the **Motion and Shape Interest Map** is calculated frame to frame, and is composed of *activated* pixels pertaining to pixels or elements of interest, *inhibited* pixels included in pixels or elements of no interest, and *neutral* pixels. In other words, this paper will show the processes related to lateral interaction and accumulative computation, whose neuronal basis has already been demonstrated in some previous papers of the same authors [2-8]. Hence, these subtasks may be implemented with a neural inspiration. Next section is devoted to describe in extensive both subtasks.

3 Accumulative computation and lateral interaction subtasks

3.1 Subtask “Working Memory Generation”

The process of obtaining the **Working Memory** consists in superimposing, just as done with superimposed transparencies, the image segmented in grey level bands of the current frame (at time instant *t*), that is to say, **Grey Level Bands** (dynamic role of inference “Segmentation in Grey Level Bands”) with image **Motion and Shape Interest Map** (dynamic role of inference “Motion and Shape Features Integration”) constructed at the previous frame (at time instant *t*-1*)*. Thus, in the **Working Memory**, at time instant *t* all scene elements associated to pixels or elements of interest will appear. This process may be observed in more detail in figure 4, where in image **Motion and Shape Interest Map** a black pixel means *neutral*, a grey pixel means *inhibited* and a white pixel means *activated.*

Coming from the image segmented in bands, some processes are performed in parallel for each band. So, there will be as many images as number of bands the image has been segmented in. These images, $B_i(x,y,t)$, are binary: 1 if the brightness of the pixel belongs to band *i*, and 0 if it does not belong to band *i.*

For each one of the binary images, $B_i(x,y,t)$, a process is carried out to fill the shapes of any band that includes any value *activated* in the classified **Motion and Shape Interest Map**, *MS(x,y,t).* For it, initially $S_i\,(x,y,t)$, the value of the filled shape at band *i* for coordinate *(x,y)* at time instant *t*, is obtained as explained next. It is assigned the value of the band if the corresponding coordinate in the **Motion and Shape Interest Map** is activated:

$$S_i(x,y,t) = \begin{cases} i, & if\ B_i(x,y,t) \equiv 1\, and\ \ MS(x,y,t) \equiv activated \\ 0, & otherwise \end{cases}$$

Afterwards, by means of lateral interaction, the same value *i* is assigned to all positions connected to $B_i(x,y,t)$, - using a connectivity of 8 pixels -, whenever its value in the **Motion and Shape Interest Map** does not indicate that it belongs to a non valid shape.

$$\forall(\alpha,\beta) \in [x \pm 1, y \pm 1], \quad S_i(\alpha,\beta,t) = \begin{cases} i, & if\ B_i(\alpha,\beta,t) \equiv 1\, and\ \ MS(\alpha,\beta,t) \neq inhibited \\ 0, & otherwise \end{cases}$$

Finally, to generate the **Working Memory**, all obtained images are summed up, as in:

$$W(x,y,t) = \sum_{i=1}^{n} S_i(x,y,t)$$

Notice that, in a given time instant *t*, scene elements whose shape features do not correspond to those defined by the observer may appear in the **Working Memory**. This may be due to the fact that these elements have not yet been labelled. But, if the features of these elements do not correspond to those indicated as interesting by the observer, these elements at t+1 will appear as *inhibited* in the **Motion and Shape Interest Map**. This way, at t+1 these elements will disappear from the **Working Memory**. In order to only obtain elements that really fulfil features defined by the observer, some accumulative computation mechanisms are introduced, as explained in section 3.2.

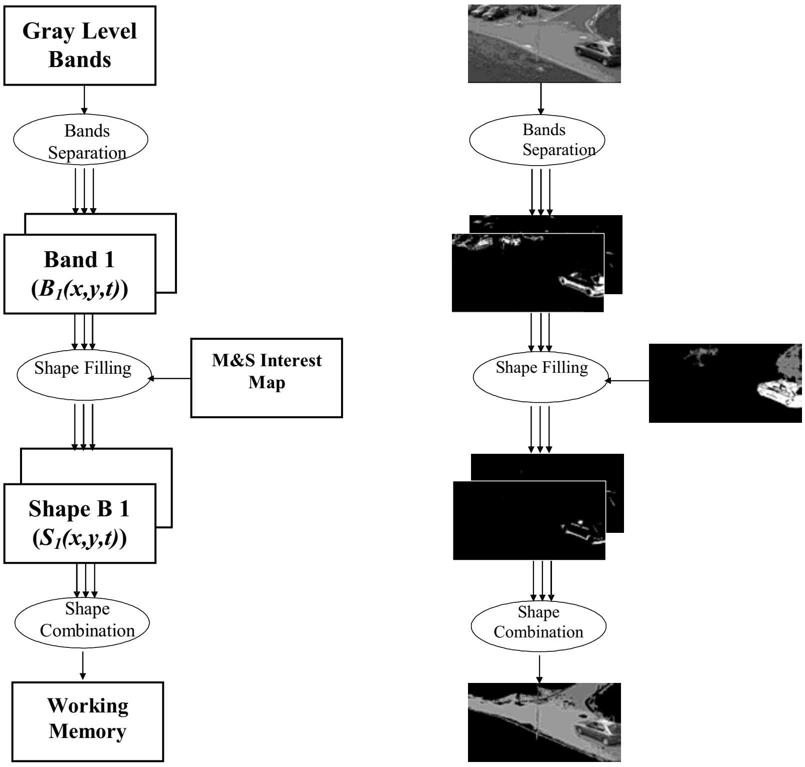

Fig. 4. Subtask “Working Memory Generation”

3.2 Subtask “Thresholded Permanency Calculation”

Subtask “Thresholded Permanency Calculation” uses as input the **Working Memory** and gets as output the final **Dynamic Attention Map**. This subtask firstly calculates the so called permanency value associated to each pixel. If the value of the **Working Memory** is activated, the charge value is incremented at each image frame by a value of δc up to a maximum. If this is not the case, the charge value is decremented by a value of δd down to a minimum. In a second step, the charge values are thresholded to get the **Dynamic Attention Map**. Thus, the **Dynamic Attention Map** will only have activated those pixels which have been *activated* in the **Working Memory** during various successive time instants, namely “threshold / charge” (see figure 5). This is shown by means of the following formulas:

$$Q(x,y,t)=\begin{cases}\min\left(Q(x,y,t-\Delta t)+\delta c, Q_{\max}\right), & \mathit{if}\ W(x,y,t)\equiv 1\\ \max\left(Q(x,y,t-\Delta t)-\delta d, Q_{\min}\right), & \mathit{if}\ W(x,y,t)\equiv 0\end{cases}$$

$$A(x,y,t)=\begin{cases}1, & \mathit{if}\ Q(x,y,t)\geq\theta\\ 0, & \mathit{otherwise}\end{cases}$$

where δc is the charge constant, δd is the discharge constant, Q_{max} is the maximum charge value, Q_{min} is the minimum charge value, θ is the threshold value and *A(x,y,t)* is the value of the **Dynamic Attention Map** at coordinate *(x,y)* at time instant *t*.

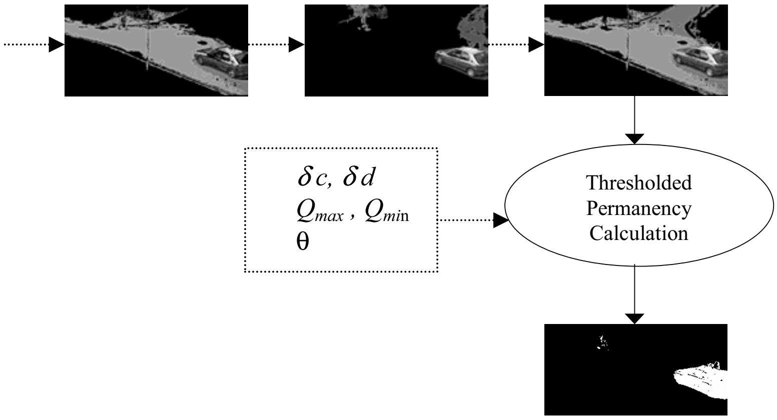

Fig.5. Subtask “Thresholded Permanency Calculation”

The accumulative computation process, followed by the threshold stage, enables to maintain active in a stable way a set of pixels which pertain to a group of scene elements of interest to the observer. Thus, the state of this “memory” defines the attention focus of the system, and is the input to the observer’s system. The observer will modify the parameters that configure the mechanisms of extraction and selection of shapes and/or pixels of interest.

4 Conclusions

A model of dynamic visual attention capable of segmenting objects in a scene has been introduced in this paper. The model enables focusing the attention at each moment at shapes that possess certain characteristics and eliminating shapes that are of no interest. The features used are related to motion and shape of the elements present in the dynamic scene. The model may be used to observe real environments indefinitely in time.

In this paper we have highlighted those subtasks of the generic “Dynamic Attention Map Generation” task related to plausible biologically inspired methods. These mechanisms, namely accumulative computation and lateral interaction, have so far been proven to be inherently neuronal. The subtasks which incorporate these biologically plausible capacities, called "Working Memory Generation" and "Thresholded Permanency Calculation", because of the fact of making use of accumulative computation and lateral interaction processes, enable maintaining a stable system of dynamic visual attention.

References

1. Eriksen, C. W., St. James, J. D.: Visual attention within and around the field of focal attention: A zoom lens model. Perception and Psychophysics 40 (1986) 225-240
2. Fernández, M.A., Mira, J.: Permanence memory: A system for real time motion analysis in image sequences. IAPR Workshop on Machine Vision Applications, MVA'92 (1992) 249-252
3. Fernández, M.A., Mira, J., López, M.T., Alvarez, J.R., Manjarrés, A., Barro, S.: Local accumulation of persistent activity at synaptic level: Application to motion analysis. In: Mira, J., Sandoval, F. (eds.): From Natural to Artificial Neural Computation, IWANN'95, LNCS 930. Springer-Verlag (1995) 137-143
4. Fernández, M.A.: Una arquitectura neuronal para la detección de blancos móviles. Unpublished Ph.D. dissertation (1995)
5. Fernández-Caballero, A., Mira, J., Fernández, M.A., López, M.T.: Segmentation from motion of non-rigid objects by neuronal lateral interaction. Pattern Recognition Letters 22:14 (2001) 1517-1524
6. Fernández-Caballero, A., Mira, J., Delgado, A.E., Fernández, M.A.: Lateral interaction in accumulative computation: A model for motion detection. Neurocomputing 50C (2003) 341-364
7. Fernández-Caballero, A., Fernández, M.A., Mira, J., Delgado, A.E.: Spatio-temporal shape building from image sequences using lateral interaction in accumulative computation. Pattern Recognition 36:5 (2003) 1131-1142
8. Fernández-Caballero, A., Mira, J., Férnandez, M.A., Delgado, A.E.: On motion detection through a multi-layer neural network architecture. Neural Networks 16:2 (2003) 205-222
9. Koch, C., & Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology 4 (1985) 219-227.
10. Itti, L., Koch, C. Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (1998) 1254-1259
11. Moore, C. M., Yantis, S., Vaughan, B.: Object-based visual selection: Evidence from perceptual completion. Psychological Science 9 (1998) 104-110
12. Posner, M.I., Snyder, C.R.R., Davidson, B.J.: Attention and the detection of signals. Journal of Experimental Psychology: General 109 (1980) 160-174
13. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cognitive Psychology 12 (1980) 97-136
14. Vecera, S.P.: Toward a biased competition account of object-based segregation and attention. Brain and Mind 1 (2000) 353-385
15. Wolfe, J.M.: Guided Search 2.0.: A revised model of visual search. Psychonomic Bulletin & Review 1 (1994) 202-238