Pattern Recognition Letters 22 (2001) 1517–1524
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Segmentation from motion of non-rigid objects by neuronal lateral interaction

A. Fernandez-Caballero [a,*], J. Mira [b], M.A. Fernandez [a], M.T. Lopez [a]

[a] *Departamento de Informatica, Universidad de Castilla-La Mancha, Campus Universitario, 02071 Albacete, Spain*
[b] *Departamento de Inteligencia Artificial, UNED, c/ Senda del Rey, s/n 28040 Madrid, Spain*

Received 3 July 2000; received in revised form 31 May 2001

Abstract

The problem we are stating is the discrimination of non-rigid objects capable of holding our attention in a scene. Motion allows gradually obtaining all moving objects shapes. We introduce an algorithm that fuses spots obtained by means of neuronal lateral interaction in accumulative computation.

Keywords: Segmentation; Non-rigid objects; Motion; Lateral interaction; Neural networks

1. Problem statement

Obtaining an object's shape from motion in a sequence of images is an arduous task in computer vision. Information of its structure may be got using a so-called structure-from-motion algorithm. Structure-from-motion algorithms are based on the information of temporary disparity using the change in the position of the moving object or of the observer in two or more successive frames.

A lot of approaches on non-rigid object segmentation have been investigated so far (e.g. Kass et al., 1988; Terzopoulos and Witkin, 1988; Pentland, 1990; Horowitz and Pentland, 1991; Kervrann and Heitz, 1994; Maurizot et al., 1995). One classical approach is that of Ullman (1979). An orthographic imaging model is used to estimate the structure and motion of a rigid moving object. The use of perspective transformations introduce non-linearities and, consequently, the complexity of the problem is substantially increased (Aggarwal and Nandhakumar, 1988). Roach and Aggarwal (1980) developed an algorithm that computes structure and motion by considering the perspective transform of the images.

The described methods are able to reduce noise using a great number of matched points in each image. An alternative way is simply to use a greater number of images. The biggest criticism to most algorithms is that they assume that there is only one object in motion. This assumption implies that a scene that contains multiple moving objects has to be segmented in its constituent objects before being able to use any algorithm.

* Corresponding author. Tel.: +34-967-59-92-00; fax: +34-967-59-92-24.
E-mail address: caballer@info-ab.uclm.es (A. Fernandez-Caballero).

PII: S0167-8655(01)00105-2

The problem that we are stating is the discrimination of a group of non-rigid objects able to maintain our attention in a scene. These objects are discovered starting from the motion of any of their parts. In an indefinite succession of images, motion allows to obtain the shape of the moving elements.

Somehow, the method can be generically included in the image difference models. More concretely, it is bound to the generic behaviour of the permanency memories (Fernandez and Mira, 1992). Specifically, we will say that the observer is unable to discern any object unless it starts moving. Therefore, an image containing only static elements does not offer any interesting information. In other words, the system only acts on those image pixels where some change in the grey level is detected between two consecutive frames.

The method is not limited to the observation, detection and pursuit of a single object in the scene, but rather it is able to segment all objects that offer some kind of motion.

2. Neuronal lateral interaction

A generic model based on a neural architecture is presented. The proposed model is based on an accumulative computation function (Fernandez and Mira, 1992; Fernandez et al., 1995), followed by a set of co-operating lateral interaction processes. These are performed on a functional receptive field organised as centre-periphery over linear expansions of their input spaces (Mira et al., 1993; Mira et al., 1995; Moreno et al., 1969).

A lateral interaction model (Mira, 1993; Mira et al., 1996) consists of a layer of modules of the same type with local connectivity, such that the response of a given module does not only depend on its own inputs, but also on the inputs and outputs of the module's neighbours. From a computational point of view, the aim of the lateral interaction nets is to partition the input space into three regions: centre, periphery and excluded. The following steps have to be done: (a) a processing over the central region, (b) a processing over the feedback of the periphery zone, (c) a comparison of the results of these operations and a local decision generation, and (d) a distribution over the output space.

We also incorporate the notion of double time-scale present at sub-cellular micro-computation (Fernandez et al., 1995). So, the following properties are applicable to the model: (a) a local convergent process around each element, (b) a semi-autonomous functioning, with each element capable of spatio-temporal accumulation of local inputs in time-scale T, and conditional discharge, and (c) an attenuated transmission of these accumulations of persistent coincidences towards the periphery that integrates at global time-scale t.

Therefore we are in front of two different time-scales: (1) the local time T, and (2) the global time $t(t = nT)$. Global time is applicable to steps (a) and (d) of our neuronal lateral interaction model, whereas steps (b) and (c) use local time-scale T.

3. Algorithms

Firstly, at step (a), a general accumulative computation present at global time-scale is performed. Each element (x, y) is capable of processing motion from input grey level value $\mathrm{IN}(x, y, t)$ and its proper charge value. Let $\mathrm{GLS}(k, x, y, t)$ be the presence or absence of grey level k at element (x, y) at time t,

$$\mathrm{GLS}(k, x, y, t) = \begin{cases} -1 & \text{if } \mathrm{IN}(x, y, t) \neq k \\ 1 & \text{otherwise} \end{cases} \quad \forall k \in [0, 255].$$

The accumulative computation equation may be formulated as

$$C(k,x,y,t) = \begin{cases} l_{\mathrm{dis}} & \text{if } \mathrm{GLS}(k,x,y,t) = -1, \\ l_{\mathrm{sat}} & \text{if } \mathrm{GLS}(k,x,y,t) = 1 \text{ and } \mathrm{GLS}(k,x,y,t-\Delta t) = -1, \\ \max(C(k,x,y,t-\Delta t) - v_{\mathrm{acc}}, l_{\mathrm{dis}}) & \\ \qquad \text{if } \mathrm{GLS}(k,x,y,t) = 1 \text{ and } \mathrm{GLS}(k,x,y,t-\Delta t) = 1, & \end{cases}$$

where $C(k,x,y,t)$ is the charge value at element (x,y) for grey level k, l_{dis} is the discharge or minimum value, l_{sat} is the saturation or maximum value, and v_{acc} is the accumulative computation discharge value due to motion detection.

Note that Δt determines the sequence frame rate and is given by the capacity of the model's implementation to process one input image.

The accumulative computation equation may be explained the following way:

(a) element (x,y) at grey level k is completely discharged if pixel $\mathrm{IN}(x,y,t)$ is not at grey level k;

(b) element (x,y) at grey level k is set to the complete charge (saturation) if pixel $\mathrm{IN}(x,y,t)$ has changed to grey level k from instant $t-\Delta t$ to instant t;

(c) element (x,y) at grey level k is partially discharged by a value of v_{acc} if pixel $\mathrm{IN}(x,y,t)$ is at grey level k and has not changed since $t-\Delta t$.

The input to steps (b) and (c) is obtained by taking the maximum value of all charge values corresponding to all possible grey levels.

$$C(x,y,t) = \max(C(k,x,y,t)) \quad \forall k \in [0,255].$$

Secondly, corresponding to steps (b) and (c), in local time-scale T, and in an iterative way, the charge value is homogeneously distributed among all connected charged elements ($C(x,y,T) > l_{\mathrm{dis}}$). An acceptable equation for this lateral interaction is

$$C(x,y,T) = \frac{1}{1+\delta_{x-1,y}+\delta_{x+1,y}+\delta_{x,y-1}+\delta_{x,y+1}} \Big[C(x,y,T-\Delta T) + \delta_{x-1,y}\, C(x-1,y,T-\Delta T) \\ + \delta_{x+1,y}\, C(x+1,y,T-\Delta T) + \delta_{x,y-1}\, C(x,y-1,T-\Delta T) + \delta_{x,y+1}\, C(x,y+1,T-\Delta T) \Big],$$

where

$$\delta_{\alpha,\beta} = \begin{cases} 1 & \text{if } C(\alpha,\beta,T-\Delta T) > l_{\mathrm{dis}}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the neighbours taken into account are four, namely the four surrounding elements (x,y). The homogeneous distribution has been obtained by performing a recursive calculus of the mean value of up to five charge values. The value of ΔT will determine the number of times the mean value is calculated.

Lastly, at step (d), the charge value $C(x,y,t)$ of each module (x,y) is distributed as the model's output.

4. Data and results

Firstly, we offer in this paper the results obtained for the traffic intersection sequence recorded at the Ettlinger-Tor in Karlsruhe by a stationary camera, copyright © 1998 by H.-H. Nagel, Institut für Algorithmen und Kognitive Systeme, Fakultät für Informatik, Universität Karlsruhe (TH), Postfach 6980, D-76128 Karlsruhe, Germany.

Fig. 1 shows two images of the sequence. You may observe the existence of 10 cars and one bus driving in three different directions. At the bottom of the image there is another car, but this one is still. The

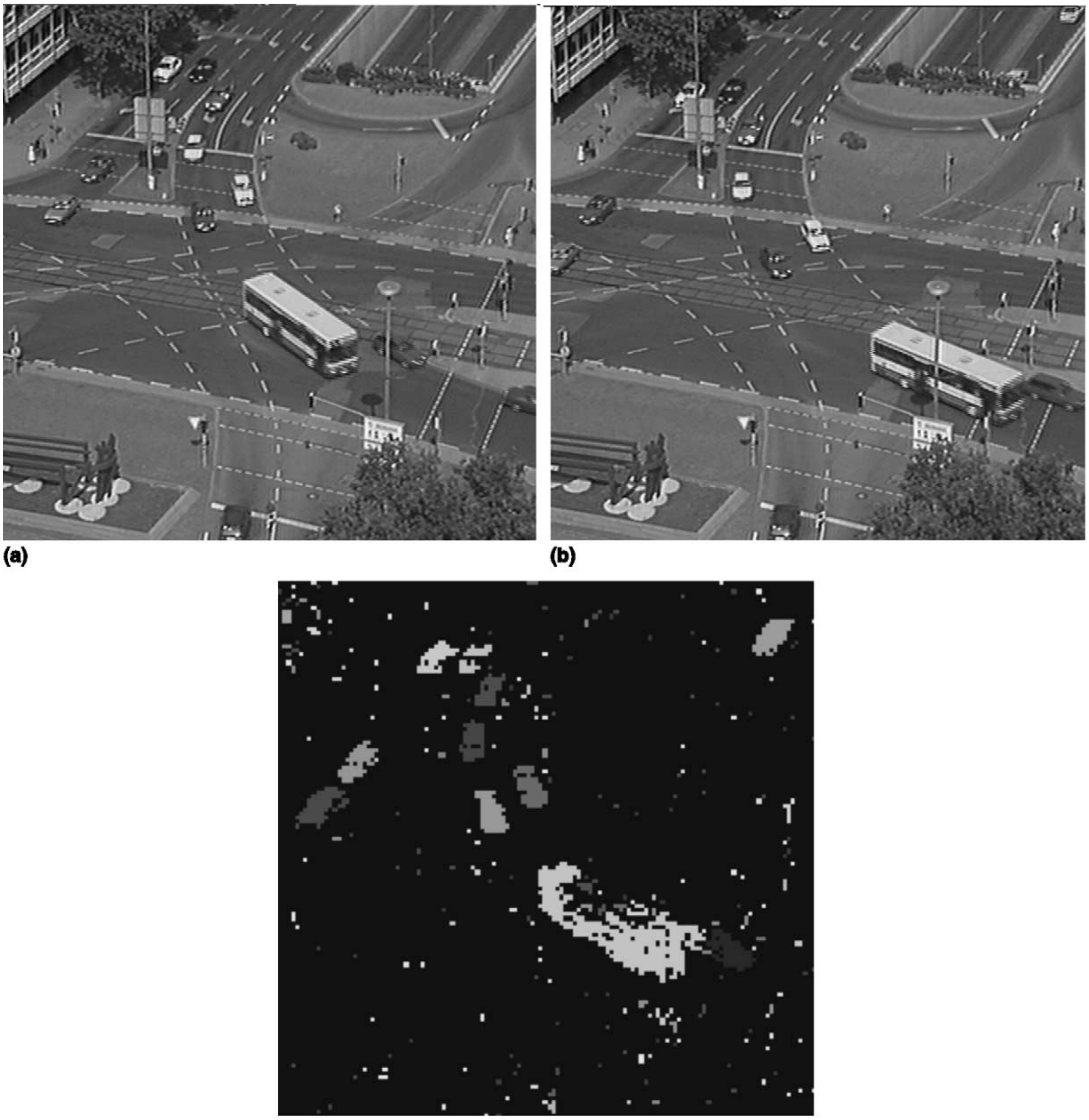

Fig. 1. Two images of the traffic intersection sequence: (a) image number 1; (b) image number 26 and (c) result of applying the neuronal lateral interaction mechanisms.

parameter values for this experiment are $\Delta t = 0.42$ s, $\Delta t = 064\,\Delta T$, $l_{\text{dis}} = 0$, $l_{\text{sat}} = 255$, and $v_{\text{acc}} = 32$. Only three frames are needed to obtain accurate segmentation results. Fig. 1(c) shows the result of applying our model to some images of the traffic intersection sequence. As you may observe, the system is perfectly capable of segmenting all the moving elements present in Fig. 1. Note that the output image has been coloured to highlight the resulting output charge values.

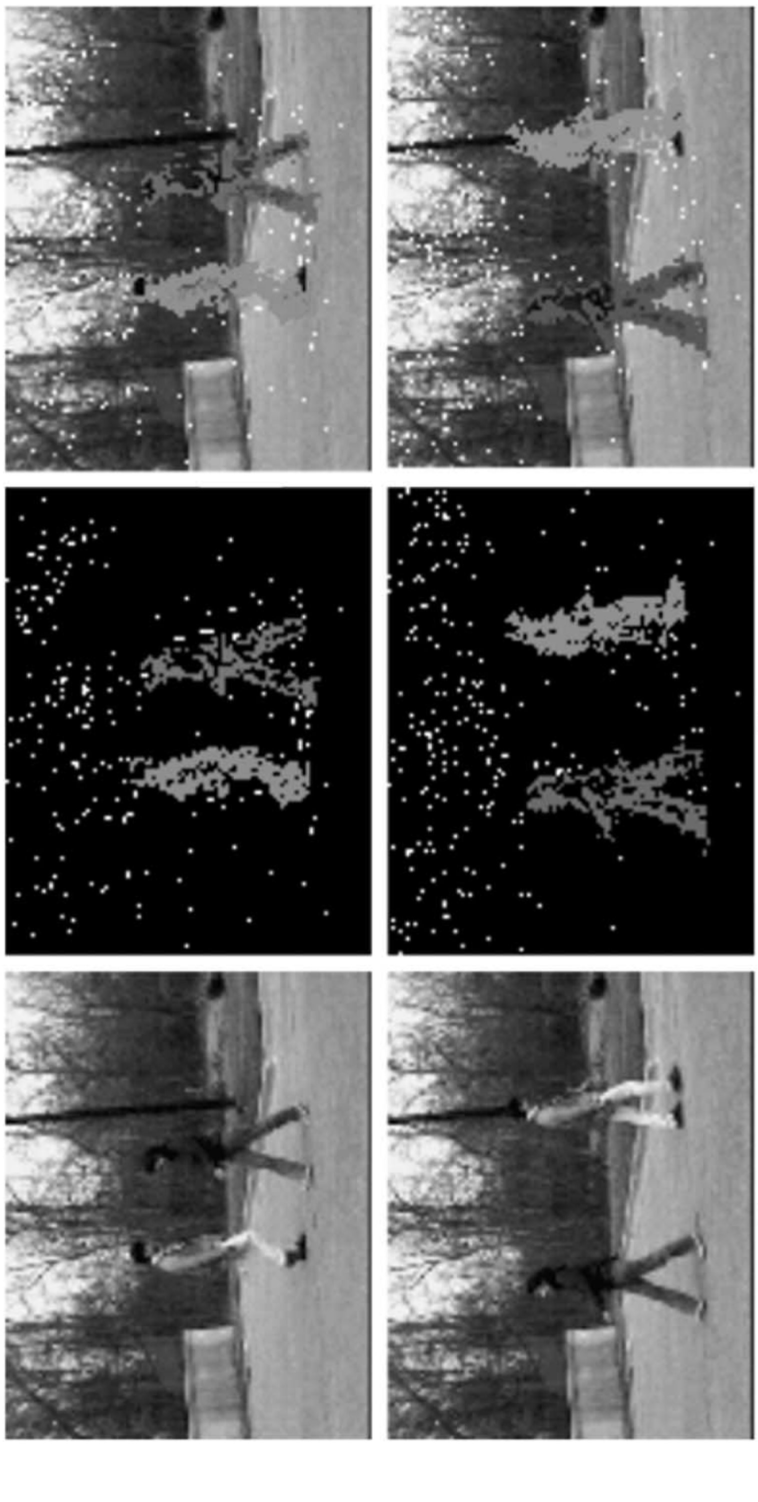

Fig. 2. Two images of the *TwoWalkNew* sequence: (a) original images; (b) result of applying the neuronal lateral interaction mechanisms and (c) image superposition.

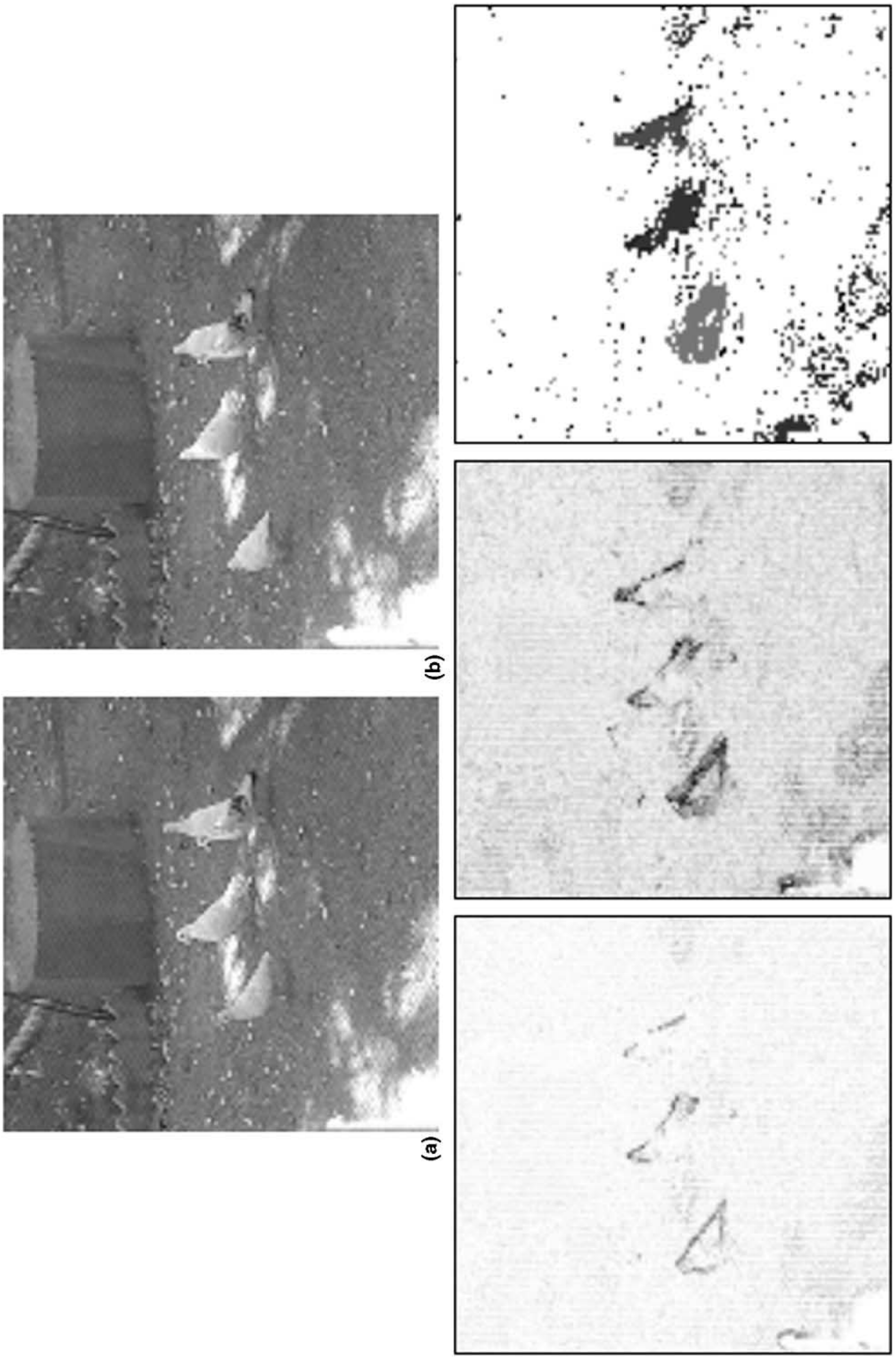

Fig. 3. Three pigeons sequence: (a) image number 1; (b) image number 5; (c) result of image difference; (d) result of cumulative image difference and (e) result of applying the neuronal lateral interaction mechanisms.

A second example is presented next (Fig. 2). Here we used the image sequence *TwoWalkNew* downloaded from University of Maryland Institute for Advanced Computer Studies, copyright © 1998 University of Maryland, College Park. This sequence, used to test the real-time visual surveillance system W^4 (Haritaoglu et al., 1998) shows two people walking trough a scene. Column (a) shows two images from the original sequence, column (b) shows the result of applying our model to the sequence, and column (c) offers the superposition of the result and the input images. The parameter values for this experiment were $\Delta t = 0.35$ s, $\Delta t = 064\,\Delta T$, $l_{\mathrm{dis}} = 0$, $l_{\mathrm{sat}} = 255$, and $v_{\mathrm{acc}} = 108$. Note again that our model perfectly detects the moving non-rigid objects in this very simple manner.

5. Conclusions

A model based on a neural architecture close to biology has been proposed in this paper. A simple algorithm of lateral interaction in accumulative computation is capable of segmenting all rigid and non-rigid objects in an indefinite sequence of images in a robust and coherent manner.

Our method may be compared to background subtraction or frame difference algorithms in the way motion is detected. Then, a region growing technique is performed to define moving objects. Fig. 3 shows that our method (Fig. 3(e)) is much stronger than simple image difference (Fig. 3(c)), and even cumulative image difference (Fig. 3(d)). Compared to both algorithms, our lateral interaction in accumulative computation model offers more accurate and less noisy results.

In contrast to similar approaches, no complex image pre-processing has to be performed, no reference image has to be offered to our model, and no high-level knowledge has to be inferred to obtain accurate results.

We also have to highlight that our proposed model has no limitation in the number of non-rigid objects to differentiate. Our system facilitates object classification by taking advantage of the object charge value, common to all pixels of a same moving element. This way, all moving objects are clearly segmented. Thanks to this fact, any higher-level operation will decrease in difficulty.

We conclude affirming that the proposed neuronal lateral interaction in accumulative computation mechanisms offer an excellent tool for image segmentation as a first approach to pattern recognition. Currently, we are studying the usefulness of our algorithms to very different real world applications such as traffic monitoring, people surveillance, and medical imaging.

Acknowledgements

Traffic intersection sequence at the Ettlinger-Tor in Karlsruhe, courtesy of Universität Karlsruhe, Fakultät für Informatik, Institut für Algorithmen und Kognitive Systeme, Kognitive Systeme (H.-H. Nagel und Mitarbeiter). *ToWalkNew* sequence, courtesy of University of Maryland Institute for Advanced Computer Studies, College Park.

References

Aggarwal, J.K., Nandhakumar, N., 1988. On the computation of motion from sequences of images – a review. Proc. IEEE 76 (8), 917–935.

Fernandez, M.A., Mira, J., 1992. Permanence memory: a system for real time motion analysis in image sequences. In: IAPR Workshop on Machine Vision Applications MVA'92, pp. 249–252.

Fernandez, M.A., Mira, J., Lopez, M.T., Alvarez, J.R., Manjarres, A., Barro, S., 1995. 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. Lecture Notes in Computer Science, Vol. 930. Springer, Berlin, pp. 137–143.

Haritaoglu, I., Harwood, D., Davis, L.S., 1998. W^4: Who? When? Where? What? A real time system for detecting and tracking people. In: Proc. Second Internat. Conf. on Automatic Face and Gesture Recognition, Nara, Japan, pp. 222–227.

Horowitz, B., Pentland, A., 1991. Recovery of non-rigid motion and structure. In: Proc. Computer Vision and Pattern Recognition Conf., Hawaii, pp. 325–330.

Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: active contour models. Internat. J. Comput. Vision 1 (4), 321–331.

Kervrann, C., Heitz, F., 1994. A hierarchical framework for the segmentation of deformable objects in image sequences. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Seattle, pp. 724–728.

Maurizot, M., Bouthemy, P., Delyon, B., Iouditski, A., Odobez, J.M., 1995. Determination of singular points in 2D deformable flow fields. In: Proc. 2nd IEEE Internat. Conf. on Image Processing, Washington, DC.

Mira, J., 1993. Computación neural en el camino visual. Notas de Visión y Apuntes sobre la Ingeniería del Software. Colección Estudios, 24, UCLM, pp. 175–197.

Mira, J., Delgado, A.E., Alvarez, J.R., de Madrid, A.P., Santos, M., 1993. Towards more realistic self contained models of neurons: high-order, recurrence and local learning. In: Mira, J., Cabestany, J., Prieto, A. (Eds.), New Trends in Neural Computation, IWANN'93, Lecture Notes in Computer Science, Vol. 686. Springer, Berlin, pp. 55–62.

Mira, J., Delgado, A.E., Boticario, J.G., Diez, F.J., 1995. Aspectos Básicos de la Inteligencia Artificial. Editorial Sanz y Torres, SL, Madrid.

Mira, J., Delgado, A.E., Manjarres, A., Ros, S., Alvarez, J.R., 1996. Cooperative processes at the symbolic level in cerebral dynamics: reliability and fault tolerance. In: Díaz, R.M., Mira, J. (Eds.), Brain Processes Theories and Models. MIT Press, Cambridge, MA, pp. 244–255.

Moreno, D.R, Rubio, F., Mira, J., 1969. Aplicación de las transformaciones integrales al proceso de datos en la retina. Rev. de Automática 5, 7–17.

Pentland, A.P., 1990. Automatic extraction of deformable part models. Internat. J. Comput. Vision 4, 107–126.

Roach, J.W., Aggarwal, J.K., 1980. Determining the motion of objects from a sequence of images. IEEE Trans. Pattern Anal. Machine Intell. 2 (6), 554–562.

Terzopoulos, D., Witkin, A., 1988. Deformable models: physically based models with rigid and deformable components. IEEE Comput. Graphics Appl. 8, 41–51.

Ullman, S., 1979. The Interpretation of Visual Motion. MIT Press, Cambridge, MA.