Lateral Interaction in Accumulative Computation: Motion-Based Grouping Method

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

[1] Universidad de Castilla-La Mancha, E.P.S.A., 02071 - Albacete, Spain
{caballer, miki, mlopez}@info-ab.uclm.es
[2] Universidad Nacional de Educación a Distancia,
E.T.S.I. Informática, 28040 - Madrid, Spain
{jmira, adelgado}@dia.uned.es

Abstract. To be able to understand the motion of non-rigid objects, techniques in image processing and computer vision are essential for motion analysis. Lateral interaction in accumulative computation for extracting non-rigid blobs and shapes from an image sequence has recently been presented, as well as its application to segmentation from motion. In this paper we show an architecture consisting of five layers based on spatial and temporal coherence in visual motion analysis with application to visual surveillance. The LIAC method used in general task ”spatio-temporal coherent shape building” consists in (a) spatial coherence for brightness-based image segmentation, (b) temporal coherence for motion-based pixel charge computation, (c) spatial coherence for charge-based pixel charge computation, (d) spatial coherence for charge-based blob fusion, and, (e) spatial coherence for charge-based shape fusion. In our case, temporal coherence (in accumulative computation) is understood as a measure of frame to frame motion persistency on a pixel, whilst spatial coherence (in lateral interaction) is a measure of pixel to neighbouring pixels accumulative charge comparison.

1 Introduction

There has been a great deal of research interest in motion tracking [1],[2],[3] because of its great applicability in a wide variety of applications. Vision is probable the most powerful source of information used by man to represent a monitored scene. Visual information is composed of a great deal of redundant sets of spatial and temporal data robustly and quickly processed by the brain. There has also been much work carried out on the extraction of non-rigid shapes from image sequences. In general, all papers take advantage of the fact that the image flow of a moving figure varies both spatially and temporally.

Little and Boyd [4] found it reasonable to suggest that variations in gaits are recoverable from variations in image sequences. There have been several attempts to recover characteristics of gait from image sequences. Polana and Nelson [5] characterize the temporal texture of a moving figure by summing the

M. De Gregorio et al. (Eds.): BV&AI 2005, LNCS 3704, pp. 396–405, 2005.

energy of the highest amplitude frequency and its multiples. Their more recent work [6] emphasizes the spatial distribution of energies around the moving figure. Bobick and Davis [7] introduced the Motion Energy Image (MEI), a smoothed description of the cumulative spatial distribution of motion energy in a motion sequence. Yang and Ahuja [8] segment an image frame into regions with similar motion. The algorithm identifies regions in each frame comprising the multiscale intraframe structure. Regions at all scales are then matched across frames. Affine transforms are computed for each matched region pair. The affine transform parameters for region at all scales are then used to derive a single motion field that is then segmented to identify the differently moving regions between two frames. Olson and Brill [9] propose a general purpose system for moving object detection and event recognition where moving objects are detected using change detection and tracked using first-order prediction and nearest neighbour matching.

Behind all of these papers one can guess the idea of grouping spatially andf temporally coherent image pixels into regions based on a common set of features. Coherence is defined as logical and orderly and consistent relation of parts. Spatial coherence describes the correlation between a set of features at different points in space. Temporal coherence describes the correlation or predictable relationship between those (or other) features observed at different moments in time. Spatial coherence is described as a function of distance (a measure or a metric), and is often presented as a function of correlation versus absolute distance between observation points. The same operation can be performed in time. It is well known that temporal and spatial coherence are involved in the promotion of perceptual binding.

The goal of this paper is to present our method for spatio-temporally shape building taking advantage of the inherent motion present in image sequences. In an indefinite succession of images, our motion-based algorithms allow to obtain the shape of the moving elements. Somehow, the method is bound to the generic behaviour of the permanency memories [10]. Specifically, we will say that the observer is unable to discern any object unless it starts moving. In other words, the system only acts on those image pixels where some change in the grey level is detected between two consecutive frames.

2 Lateral Interaction in Accumulative Computation (LIAC)

Lateral interaction in accumulative computation has recently been introduced [11],[12], as well as its application to segmentation from motion [13]. For it, a generic model based on a neural architecture was presented. We shall now remind of the most important characteristics of this model. The proposed model is based on accumulative computation function followed by a set of cooperating lateral interaction processes. These are performed on a functional receptive field organised as centre-periphery over non-linear and temporal expansions of their input spaces. A lateral interaction model 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 neighbors. 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) processing over the central region, (b) processing over the feedback of the periphery zone, (c) comparison of the results of these operations and a local decision generation, and, (d) distribution over the output space.

We also incorporate the notion of double time scale present at sub-cellular microcomputation. So, the following properties are applicable to the model. (a) Local convergent process around each element, (b) semiautonomous functioning, with each element capable of spatio-temporal accumulation of local inputs in time scale T, and conditional discharge, and, (c) 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, $(t = n \cdot T)$. 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 LIAC for Spatio-Temporal Coherent Shape Building

In first place, and in the following figure, the complete structure chosen as the modular computational solution to apply the model to spatio-temporal shape building is presented.

In Figure 1, five layers can be appreciated that form the architecture of the lateral interaction in accumulative computation method.

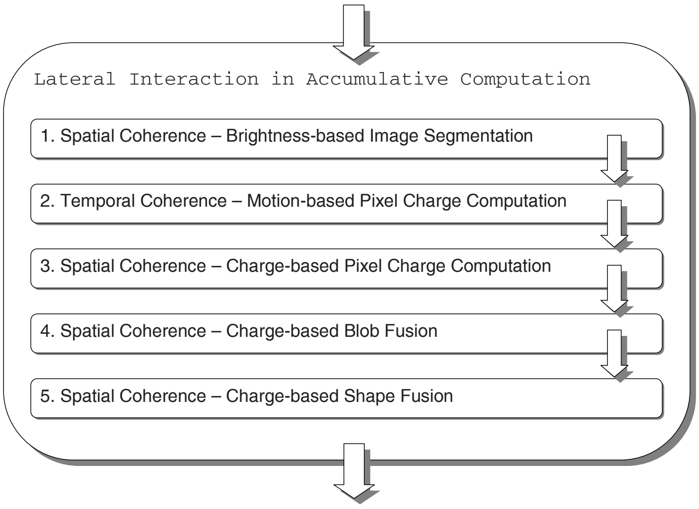
Lateral Interaction in Accumulative Computation
1. Spatial Coherence – Brightness-based Image Segmentation
2. Temporal Coherence – Motion-based Pixel Charge Computation
3. Spatial Coherence – Charge-based Pixel Charge Computation
4. Spatial Coherence – Charge-based Blob Fusion
5. Spatial Coherence – Charge-based Shape Fusion

Fig. 1. LIAC architecture for coherent shape building

Now we are going to explain the role of each of these five layers devoted to shape building. As it will be easy to appreciate, in each of these layers seeking for coherence is the main objective. In effect, layers 1, 3, 4 and 5 are based on spatial coherence, whereas layer 2 is a typical application of temporal coherence. The consistency of the LIAC method for spatio-temporal coherent shape building lays on motion-based grouping of pixels and blobs.

3.1 Spatial Coherence – Brightness-Based Segmentation

This layer covers the possibility to segment the image into a predefined group of n grey level bands just from the brightness of each input image pixel. This layer enables to smoothening the transitions among neighbouring pixels of the input image. This may be considered a first step that contributes to spatial coherence.

Let $GL(x,y,t)$ be the input grey level value at element (x,y) at time t, and let $GLS(k,x,y,t)$ be the presence or absence of grey level k at element (x,y) at time t. Then

$$GLS(k,x,y,t) = \begin{cases} 1, & \text{if } \frac{GL[x,y,t]}{GL_{max}-GL_{min}+1} + 1 = k, \forall k \in [0, n-1] \\ -1, & \text{otherwise} \end{cases} \quad (1)$$

where n is the *number of grey level bands*, and, k is a particular grey level band.

In other words, we are determining in which grey level band a certain pixel falls. So, we are not evaluating, at this level, if there is motion in a grey level band for a given pixel, but a brightness-based spatially coherent segmentation is performed. Coherence, in this case, has to be understood as the relation of belonging to a same grey level band.

It must be clear that one, and only one, of the outputs of all the detecting modules of the grey level bands can be activated at a given instant. This fact, although obvious, is of a great interest at the higher layers of the architecture, since it will avoid possible conflicts among the values offered by the different grey level bands. Indeed, only one grey level band will contain valid values.

3.2 Temporal Coherence – Motion-Based Pixel Charge Computation

This layer has been designed to obtain the permanence value $PM(k,x,y,t)$ [10], [11] on a decomposition in grey level bands basis. We will have n sub-layers and each one of them will memorise the value of the accumulative computation present at global time scale t for each element. Lateral interaction in this layer is thought to reactivate the permanence charge of those elements partially loaded and that are directly or indirectly connected to maximally charged elements. The permanence charge of each element will be offered as the input of the following layer.

Firstly, at global time scale t, permanence memory charge or discharge due to motion detection is performed. This information, given as input from the previous layer, is associated to sub-layer k of layer 1 (grey level band k). The accumulative computation equation may be formulated as

$$PM(k,x,y,t) = \begin{cases} l_{dis}, \text{if } GLS(k,x,y,t) = -1 \\ l_{sat}, \text{ if } GLS(k,x,y,t) = 1 and GLS(k,x,y,t-\triangle t) = -1 \\ \max(PM(k,x,y,t-\triangle t) - d_v, l_{dis}), \\ \qquad \text{if } GLS(k,x,y,t) = 1 and GLS(k,x,y,t-\triangle t) = 1 \end{cases} \tag{2}$$

where l_{dis} is the discharge or $Minimum\ permanence\ value$, l_{sat} is the saturation or $Maximum\ permanence\ value$, and, d_v is the *Discharge value due to motion detection*.

Note that t determines the sequence frame rate and is given by the capacity of the model's implementation to process one input image. At each element (x, y) we are in front of three possibilities: (1) The sub-layer does not correspond to the grey level band of the image pixel. The permanence value is discharged down to value l_{dis}. (2) The sub-layer corresponds to the grey level band of the image pixel at time instant t, and it didn't correspond to the grey level band at the previous instant $t - \triangle t$. The permanence value is loaded to the maximum of saturation l_{sat}. (3) The sub-layer corresponds to the grey level band of the image pixel at time instant t, and it also corresponded to the grey level band at the instant $t - \triangle t$. The permanence value is discharged by a value d_v (discharge value due to motion detection); of course, the permanence value cannot get off a minimum value l_{dis} . The discharge of a pixel by a quantity of d_v is the way to stop maintaining attention to a pixel of the image that had captured our interest in the past. Notice that we really are in front of a temporal coherence mechanism, where coherence depends on the comparison between the grey level bands of each pixel at two consecutive time instants (two sucessive frames).

3.3 Spatial Coherence – Charge-Based Pixel Charge Computation

Obviously, if a pixel is not directly or indirectly bound by means of lateral interaction mechanisms to a maximally charged pixel (l_{sat}), it goes down to the total discharge with time. That is why, secondly, an extra charge r_v (*Recharge value due to neighbouring*) is added to the permanence memory in those image pixels that receive a stimulus from a maximally charged element almost l_1 pixels far away in any of four directions. This recharge can only happen one time, and provided that none neighbour element up to the maximally charged element is discharged. l_1 is called $Number\ of\ neighbours\ in\ accumulative\ computation$. This recharge mechanism allows maintaining attention on those pixels directly or indirectly connected to maximally charge pixels. This mechanism is even able to reinforce the permanence memory value if the $r_v > d_v$.

$$PM(k,x,y,t) = min(PM(k,x,y,t) + \epsilon \cdot r_v, l_{sat}) \tag{3}$$

where

$$\epsilon = \begin{cases} 1, \text{ if } \exists(i \leq l_1)|\forall(1 \leq j \leq i) \\ \quad ((PM(k,x+i,y,t)) = l_{sat} \bigcap (PM(k,x+j,y,t)) \neq l_{dis} \bigcup \\ \quad (PM(k,x-i,y,t)) = l_{sat} \bigcap (PM(k,x-j,y,t)) \neq l_{dis} \bigcup \\ \quad (PM(k,x,y+i,t)) = l_{sat} \bigcap (PM(k,x,y+j,t)) \neq l_{dis} \bigcup \\ \quad (PM(k,x,y-i,t)) = l_{sat} \bigcap (PM(k,x,y-j,t)) \neq l_{dis}) \\ \\ 0, \text{otherwise} \end{cases} \tag{4}$$

Lastly, back at global time scale t, the permanence value at each pixel (x, y) is threshold (θ_1) and sent to the next layer.

$$PM(k,x,y,t) = \begin{cases} PM(k,x,y,t), & \text{if } PM(k,x,y,t) > \theta_1 \\ \theta_1, & \text{otherwise} \end{cases} \tag{5}$$

In order to explain the central idea of this layer, we will say that the activation toward the lateral modular structures (up, down, right and left) is again based on coherence, this time spatial coherence. Spatial coherence is related to the permanence memory values of neighbouring pixels up to a distance of l_1. The algorithm looks for coherent permanency value paths.

Now here are the basic ideas underlying lateral interaction at this layer. (1) All modular structures with maximum permanence value l_{sat} (saturated) output the charge toward the neighbours. (2) All modular structures with a not saturated charge value, and that have been activated from some neighbour, allow passing this information through them (they behave as transparent structures to the charge passing). (3) The modular structures with minimum permanence value l_{dis} (discharged) stop the passing of the charge information toward the neighbours (they behave as opaque structures). Therefore, we are in front of an explosion of lateral activation beginning at the structures with permanence memory set at l_{sat}, and that spreads lineally toward all the addresses, until a structure appears in the pathway with a discharged permanence memory.

3.4 Spatial Coherence – Charge-Based Blob Fusion

Layer 4 is also formed of n sub-layers, where, by means of lateral interaction, charge redistribution among all connected neighbours in a surrounding window of $l_2 * l_2$ pixels that hold a minimum charge, is performed. Besides distributing the charge $C(k,x,y,t)$ in grey level bands, at this level, the charge due to the motion of the background is also diluted. The new charge obtained in this layer is offered as an output toward layer 5. Starting from the values of the permanence memory in each pixel on a grey level band basis, we will see how it is possible to obtain all the parts of an object (blobs) in movement. A blob concretely means the union of pixels that are together and in a same grey level band. The discrimination of each one of the blobs is equally obtained by lateral co-operation mechanisms. In case of layer 4, the charge will be homogenised among all the pixels that pertain to the same grey level band and that are directly or indirectly united to each other, providing a means towards spatial coherence.

This way, a double objective will be obtained: (1) Diluting the charge due to the false image background motion along the other pixels of the background. This way, there should be no presence of motion characteristic of the background, but we will rather keep motion of the objects present in the scene. (2) Obtaining a parameter common to all the pixels of the blobs in a surrounding window of $l_2 * l_2$ pixels with a same grey level band. Initially, at global time scale t, the charge value at each pixel (x, y) and at each sub-layer k is given the value of the permanence value from the previous layer. After-wards, at local time scale T, provided that the neighbour input charge values are high enough, the centre element (x, y) calculates the mean of its value and the partially charged neighbours in a surrounding window of $l_2 * l_2$ pixels. l_2 is denominated *Number of neighbours in charge redistribution.*

$$C(x,y,T) = \frac{C(k,x,y,T-\triangle T) + \sum_{i=-l_2}^{l_2} \sum_{i=-l_2}^{l_2} \delta_{x+i,y+j} \cdot C(k,x+i,y+j,T-\triangle T)}{1 + \sum_{i=-l_3}^{l_3} \delta_{x+i,y+j}}, \quad \forall (i,j) \neq (0,0) \tag{6}$$

where

$$\delta_{\alpha,\beta} = \begin{cases} 1, & \text{if } C(k,\alpha,\beta,T-\triangle T) > l_{dis} \\ 0, & otherwise \end{cases} \tag{7}$$

Again at global time scale t, the charge value at each pixel (x, y) is threshold (θ_2) and sent to the next layer.

$$C(k,x,y,t) = \begin{cases} C(k,x,y,t), & \text{if } C(k,x,y,t) > \theta_2 \\ \theta_2, & otherwise \end{cases} \tag{8}$$

3.5 Spatial Coherence – Charge-Based Shape Fusion

In each element of layer 5, we have an input from each corresponding element of the n sub-layers of layer 4. This layer has as purpose the fusion into uniform shapes of the objects in a surrounding window of $l_3 * l_3$ pixels. That is why it takes the input charges of each one of the grey level bands and performs a fusion of these values, obtaining uniform parts of all the moving objects of the original image. Its output is a set of shapes $S(x, y, t)$. Up to now attention has been captured on any moving objects in the scene by means of co-operative calculation mechanisms in all grey level bands. Motion due to background has also been eliminated. It is now necessary to fix as a new objective to clearly distinguish the motion of the different objects. This discrimination is obtained equally by lateral cooperation mechanisms. Nevertheless, now we will no longer work with sub-layers, but rather all information of the n sub-layers of layer 4 end up in a single layer. In layer 5, we will homogenise the charge values among all the pixels that contain some charge value superior to a minimum threshold and that are physically connected to each other in a radius of l_3 pixels. This is again the criteria used for spatial coherence. Firstly, the shape charge value at each pixel (x, y) is given the charge value of the maximally charged sub-layer k from the previous layer.

$$S(x, y, t) = max(C(k, x, y, t)), \forall k \in [0, 255] \tag{9}$$

At local time scale, provided that the neighbour input charge values are high enough, the centre element (x, y) calculates the mean of its value and the partially charged neighbours in a surrounding window of $l_3 * l_3$ pixels. l_3 is denominated $Number\ of\ neighbours\ in\ object\ fusion.$

$$S(x,y,T) = \frac{S(x,y,T-\triangle T) + \sum_{i=-l_3}^{l_3} \sum_{i=-l_3}^{l_3} \delta_{x+i,y+j} \cdot S(x+i, y+j, T-\triangle T)}{1 + \sum_{i=-l_3}^{l_3} \delta_{x+i,y+j}},$$

$$\forall (i,j) \neq (0,0) \tag{10}$$

where

$$\delta_{\alpha,\beta} = \begin{cases} 1, & \text{if } S(k, \alpha, \beta, T - \triangle T) > l_{dis} \\ 0, & otherwise \end{cases} \tag{11}$$

Back to global time scale t, the shape charge value at each pixel (x, y) is again threshold (θ_3).

$$S(x, y, t) = \begin{cases} S(x, y, t), & \text{if } S(k, x, y, t) > \theta_3 \\ \theta_3, & otherwise \end{cases} \tag{12}$$

4 Data and Results

In this section we offer some results of applying our LIAC method in visual surveillance to 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.

Figure 2 shows two images of the sequence. You may observe the existence of ten 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 parameter values for this experiment are $\triangle t = 0.42$ seconds, $\triangle t = 64 * T$, l_{dis}=0, l_{sat}=255 and d_v=32. Only three frames are needed to obtain accurate segmentation results. Figure 2c 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 on Figure 2. Note that the grey levels of the output image are consistent with the charge values common to the shapes obtained.

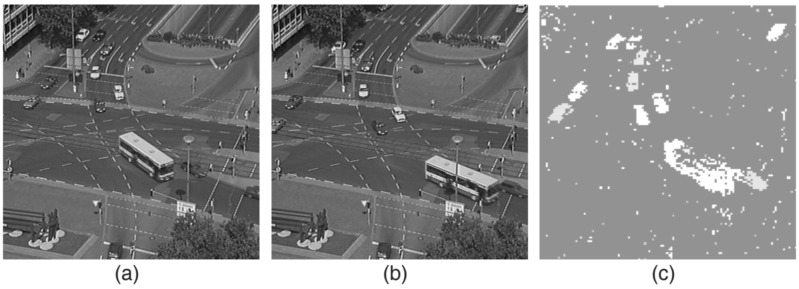

Fig. 2. Two images of the traffic intersection sequence. (a) Image number 1. (b) Image number 26. (c) Result of applying the lateral interaction mechanisms [13].

5 Conclusions

A simple algorithm of lateral interaction in accumulative computation, which is capable of segmenting all rigid and non-rigid objects in an indefinite sequence of images in a robust and coherent manner, with application to visual surveillance, has been proposed in this paper. Our method may be compared to background subtraction or frame difference algorithms in the way motion is detected. But, the main difference is that we look for spatial coherence through segmentation in grey level bands. Then, a region growing technique, based on spatio-temporal coherence of charge values assigned to image pixels, is performed to define moving objects. In contrast to similar approaches, no complex image preprocessing 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. Our model is a 2-D approach to motion estimation. In these kinds of approaches, motion estimates are obtained from 2-D motion of intensity patterns. In these methods there is a general restriction: the intensity of the image along the motion trajectory must be constant, that is to say, any change through time in the intensity of a pixel is only due to motion. This restriction does not affect our model at all. This way, our algorithms are prepared to work with lots of situations of the real visual surveillance world, where changes in illumination are of a real importance.

The gradient-based estimates have become the main approach in the applications of computer vision. These methods are computationally efficient and satisfactory motion estimates of the motion field are obtained. The disadvantages common to all methods based on the gradient also arise from the logical changes in illumination.

Obviously, a way of solving the former limitations of gradient-based methods is to consider image regions instead of pixels. In general, these methods are less sensitive to noise than gradient-based methods. Our particular approach takes advantage of this fact and uses all available neighbourhood state information as well as the proper motion information. On the other hand, our method is not affected by the greatest disadvantage of region-based methods. Our model does not depend on the pattern of translation motion. In effect, in region-based methods, regions have to remain quite small so that the translation pattern remains valid. 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.

Acknowledgements

Traffic intersection sequence at the Ettlinger-Tor in Karlsruhe, courtesy of Universität Karlsruhe, Fakultät für Informatik, Institut für Algorithmen und Kognitive Systeme. This work is supported in part by the Spanish CICYT TIN2004-07661-C02-01 and TIN2004-07661-C02-02 grants.

References

1. Huang, T.S.: Image Sequence Analysis. Springer-Verlag (1983)
2. Aggarwal, J.K., Nandhakumar, N.: On the computation of motion from sequences of images - A review. Proceedings of the IEEE **76**:8 (1988)
3. Wang, J., Huang, T.S., Ahuja, N.: Motion and Structure from Image Sequences. Springer-Verlag (1993)
4. Little, J.J., Boyd, J.E.: Recognizing people by their gait: The shape of motion. Videre: Journal of Computer Vision Research **1**:2 (1998) 2-32
5. Polana, R., Nelson, R.: Detecting activities. Proceedings of the IEEE Conference on Com-puter Vision and Pattern Recognition (1993) 2–7
6. Polana, R., Nelson, R.: Recognition of nonrigid motion. Proceedings DARPA Image Understanding Workshop (1994) 1219–1224
7. Bobick, A.F., Davis, J.W.: An appearance-based representation of action. Proceedings 13th International Conference on Pattern Recognition (1996) 307–312
8. Yang, M.-H., Ahuja, N.: Extracting gestural motion trajectories. Proceedings 2nd International Conference on Automatic Face and Gesture Recognition (1998) 10–15
9. Olson, T., Brill, F.: Moving object detection and event recognition algorithms for smart cameras. Proceedings DARPA Image Understanding Workshop (1997) 159–175
10. Fernández, M.A., Fernández-Caballero, A., López, M.T., Mira, J.: Length-speed ratio (LSR) as a characteristic for moving elements real-time classification. Real-Time Imaging **9** (2003) 49–59
11. 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
12. Mira,J.,Delgado, A.E., Fernández-Caballero, A.,Fernández, M.A.: Knowledge modelling for the motion detection task: The algorithmic lateral inhibition method. Expert Systems with Applications **27**:2 (2004)169–185
13. 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