Knowledge-Based Silhouette Detection

Antonio Fernández-Caballero

Escuela Politécnica Superior de Albacete, Departamento de Informática
Universidad de Castilla-La Mancha, 02071 – Albacete, Spain

Abstract. A general-purpose neural model that challenges image understanding is presented in this paper. The model incorporates accumulative computation, lateral interaction and double time scale, and can be considered as biologically plausible. The model uses - at global time scale *t* and in form of accumulative computation - all the necessary mechanisms to detect movement from the grey level change at each pixel of the image. The information on the detected motion is useful as part of an object's shape can be obtained. On a second time scale base $T<<t$, and by means of lateral interaction of each element with its neighbours, other parts of the moving object are also considered, even when no variation in grey level is detected on these parts. After introducing the general concepts of the model denominated Lateral Interaction in Accumulative Computation, the model is applied to the problem of silhouette detection of all moving elements in an indefinite sequence of images. The model is lastly compared to the most important current knowledge on motion analysis showing, this way its suitability to most well-known problems in silhouette detection.

1 Introduction

The visual system is able to quickly calculate the movement of the objects in an environment from the light intensity variations that reach the eye. One of the solutions the visual system has taken for the calculation of the movement is the spatial extraction of local motion signals and their spatial combination in higher processing layers to calculate more complex movement types. Current available knowledge suggests the existence of several stages in motion analysis in the visual system [1]. In the first level the measures of local movement extract those motion components perpendicular to the edges of the elements present in the image. The second level combines the measures of local movement of portions of the image with the purpose of calculating a smaller number of local estimates of translation of the object. Finally, a third level integrates the local estimates of translation movement to calculate more complex non-local movements (for example, global rotations).

The problem stated is the discrimination of a set of non-rigid objects capable of holding our attention in a scene [2],[3]. These objects are detected from the motion of any of their parts. Detected in an indefinite sequence of images, motion allows obtaining the silhouettes of the moving elements. The method introduced is compared to the state-of-the art in current knowledge for silhouette detection.

Fernández-Caballero A. (2005).
Knowledge-Based Silhouette Detection.
In *Proceedings of the 5th International Workshop on Pattern Recognition in Information Systems*, pages 114-123
DOI: 10.5220/0002570001140123

2 General-Purpose LIAC Model

A generic model based on neural architecture, with recurrent and parallel computation at each specialised layer, and sequential computation between consecutive layers, is presented. The model is based on an accumulative computation function, followed by a set of co-operating lateral interaction processes performed on a functional receptive field organised as centre-periphery over linear expansions of their input spaces [4]. Double time scale is a third important cue. All these terms will be introduced by offering a view of all models incorporating these items.

2.1 Recurrent Lateral Interaction Model

In lateral interaction models [4], you have a layer of modules of the same type with local connectivity, such that the response of a given element does not only depend on its own inputs, but also on the inputs and outputs of the element'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 (C), periphery (P) and excluded (E). The following steps must be followed: (a) processing on the central region, (b) processing on the feedback data from the periphery zone, (c) comparison of the results of these operations and a local decision generation, and, (d) distribution over the output space. There are two general expressions to formulate lateral interaction: non-recurrent interaction and recurrent interaction.

Let $I(\alpha,\beta)$ be the input signal to an element situated on co-ordinates (α,β), $\Phi(x,y)$ be the output signal of an element on position (x,y), and $K(x,y;\alpha,\beta)$ be the weight factor that translates the effect of element (α,β) into element (x,y). The output of element (x,y) for the common case of recurrent interaction is:

$$\Phi(x,y) = I(x,y) + \iint_R K(\alpha,\beta;x,y)\Phi(\alpha,\beta)d\alpha d\beta \quad (1)$$

2.2 Accumulative Computation Model

Next the accumulative computation model [5] is introduced. This model basically responds to a sequential module represented by its state value. The accumulative computation process works on an input parameter and responds with an output called the module's discharge value. The state value is also called the permanence value and is generally stored in a permanence memory. The output value of element (x,y), $\Phi(x,y,t)$, is a function of the charge value $P(x,y,t)$, as shown in equation (2):

$$\Phi(x,y,t) = \begin{cases} 0, & if\ \ P(x,y,t) < \theta \\ P(x,y,t), & otherwise \end{cases} \quad (2)$$

The module calculates at each time instant t the charge value $P(x,y,t)$ as a function of the proper permanence value at $t\text{-}\Delta t$, that is to say $P(x,y,t\text{-}\Delta t)$, and as a function of the input signal $I(x,y,t)$. The permanence function is called F_p. Note that in this accumulative computation model the previous result has to fall between two established values v_{dis} and v_{sat} corresponding to the states of discharged and saturated of element *(x,y)* at instant *t.*

$$P(x,y,t)=\begin{cases} v_{sat}, & if \quad F_p[...] \geq v_{sat} \\ F_p\left[P(x,y,t-\Delta t),I(x,y,t)\right], & if \ v_{dis} < F_p[...] < v_{sat} \\ v_{dis}, & if \ v_{dis} \geq F_p[...] \end{cases} \quad (3)$$

The structure of the input and output spaces is that of FIFO memories to include time as a calculus variable. In a layer output space there is a heap of responses of these layer units. So, it works as a local and transitory memory that saves the outputs of all the neurones of the layer during some time interval. This FIFO output memory specifies the co-operating capacity of the net (its local connectivity), as the different units in recurrent neurones layers take a look on their neighbour's responses in that moment as well as in some previous moments.

2.3 Double Time Scale Model

The model also incorporates the notion of double time scale at accumulative computation level present at sub-cellular micro-computation [5]. The following properties are applicable to the model: (a) a local convergent process around each element, (b) a semiautonomous 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 the global time scale t. Therefore there are two different time scales: (a) the local time $T = n\ \Delta T$, and, (b) the global time $t = k\ \Delta t$ ($T<<t$).

2.4 Lateral Interaction in Accumulative Computation Model

Lastly, the model proposed incorporates all the general notions seen so far, grouping common terms in biology such as accumulative computation, lateral interaction and double time scale. Note that all of them have been largely studied over time. The contribution consists in restricting the model to the following characteristics:

1. Application of accumulative computation to a single central element starting from the state value of the element itself and the input coming from the previous layer.
2. Application of lateral interaction mechanisms (lateral inhibition) from close periphery formed by four elements with co-ordinates *(x-1,y)*, *(x+1,y)*, *(x,y-1)* and *(x,y+1)* on the centre *(x,y)*. (a) The interaction is of recurrent type. (b) All neighbours have the same weight in lateral interaction. (c) The total effect on an element subject to lateral interaction is linear, that is to say the different particular effects on an element are summed up. (d) Concerning the dimensionality of the model, it may be stated that lateral interaction takes place in spatio-temporal co-ordinates.
3. Global time scale *t* is used for (a) reading the input from the previous layer, (b) data processing by accumulative computation, and (c) writing the output to the following layer.
4. Local time scale $T<<t$ is used in all lateral co-operative interaction mechanisms.

These series of specific characteristics permit to rewrite equations (1) and (3) for the particular model in the following way:

1. At global time scale t:

$$C(x,y,t)=\begin{cases} v_{sat}, & if \quad F_g[\ldots]\geq v_{sat} \\ F_g[C(x,y,t-\Delta T)\ ,\ I(x,y,t)], & if \quad v_{dis}<F_g[\ldots]<v_{sat} \\ v_{dis}, & if \quad v_{dis}\geq F_g[\ldots] \end{cases}$$

2. At local time scale T:

$$C(x,y,T)=\begin{cases} v_{sat}, & if \quad F_l[\ldots]\geq v_{sat} \\ F_l\left[C(x,y,T-\Delta T)\ ,\ \sum_{\alpha=x-1}^{x+1}\sum_{\beta=y-1}^{y+1}C(\alpha,\beta,T)\right], & if \quad v_{dis}<F_l[\ldots]<v_{sat} \\ v_{dis}, & if \quad v_{dis}\geq F_l[\ldots] \end{cases}$$

$$\forall(\alpha,\beta)\neq(x,y),\ (\alpha,\beta)\in R,(a,b)\in R$$

Now, a description of the model structure as well as the generic tasks applicable to the model may be provided. Each module's charge value $C(x,y,t)$ is a function of the history of the input signals coming from at most two preceding levels and the charge values of that particular module's four direct neighbours.

3 Knowledge-Based Silhouette Detection

Next an extensive explanation of the approach introduced when confronting the motion knowledge domain from the perspective of the lateral interaction in accumulative computation model is presented, as well as the reasons that have motivated these decisions. In greater or lesser extent, a significant number of problems appear in all models related to motion analysis when looking for the silhouettes of the objects present in a scene. This is due to the very nature of motion's projection. The most relevant drawbacks in this context are detailed. (1) The aperture problem, (2) the limits of movement, (3) the occlusions, (4) the false movement due to image background, (5) the motion speed, (6) the colour in motion, and, (7) the discrimination of more than one moving object.

3.1 The Aperture Problem

A consequence of the no equivalence between the projected movement and the optic flow is that motion information is intrinsic to the image structure. In other words, it is dependent on the variation of grey levels in the image. The aperture problem appears when there is not enough variation in grey level in a studied region of the image to uniquely resolve the problem. More than one motion candidate would be equally valid in the observed data of the image. Specifically, this means that the velocity component can only be estimated with some degree of certainty in a direction perpendicular to a significant image gradient.

The aperture problem is dealt by assigning the true velocity of the whole model to those moving image elements with possible ambiguities. Thus it can be solved if at least two movement measurements of local components exist to be able to estimate the velocity of a pattern in one point. In a simple translation movement in a plane, the problem can be resolved. This is not the case, however, for general 3D or rotational movement, where the real 2D velocity varies from point to point. It is easy to understand that the problem is even increased when dealing with non-rigid objects.

The application of lateral interaction in accumulative computation is perfectly able to confront the aperture problem by softening the true 2D velocity in each point of the silhouette to obtain an object. In fact, this occurs through the necessary lateral interaction exchange mechanisms among the points that are next in the image. Lateral interaction does not try to find the speed of each image pixel. It rather obtains a unique velocity for each moving object present in the scene. The aperture problem is solved using an adequate charge calculus function in local time scale *T*.

3.2 The Limits of Movement

The inherent ambiguity of the 3D in 2D projection causes an additional complication. Where a discontinuity exists in the depth of a scene, for example, in case the objects in movement are superimposed, some points of the three-dimensional space with different movements may be projected on the same 2D point of the image plane. This causes a discontinuity in the spatial motion field. Since the region really contains a series of different movements, a model that assumes the very common supposition that every region bears one single movement will fail when modelling these inner regional discontinuities and will estimate a unique movement for this region. This unique movement will correspond to the dominant or the mean movement. Moreover, if the discontinuities in the movement were solely temporary, this would greatly affect the algorithms of obtaining the silhouettes and their subsequent tracking.

In this case, a very relevant problem appears when trying to segment an image in all its moving objects. In this case it is also imperative to consider the outreach of the lateral interaction mechanisms. In other words, it is fundamental to perfectly delimit the receptive field *R* in the formula of lateral interaction in the local time scale *T*. Notice that the perfect demarcation of the receptive field is a fundamental learning task of lateral interaction in accumulative computation model. Lastly let us emphasise that the receptive field denotes the outreach of lateral interaction, and, therefore, movement limits.

3.3 The Occlusions

An additional consequence of the projection ambiguity is that an object in movement will expose a previously covered area and vice versa. This effect is known as occlusion. In these regions correspondence does not exist among different frames and motion field is not defined. Similar situations occur when there is a change in a given scene or when new objects enter or leave the scene. Thus, motion analysis raises its complexity, since motion has to be obtained by other means in these areas.

It is necessary to emphasise that the occlusion problem is not resolved by the lateral interaction in accumulative computation model. The model is pixel oriented (point, element) and not region oriented (area, object); hence, it inherits all problems related to occlusions in pixel-oriented segmentation.

3.4 The False Movement Due to the Image Background

It is important to pay attention to the problem of the false movement due to the image background. Indeed, whenever motion is detected due to a moving object, there is always a detection of motion due to the image background. The movement of an object constitutes an invasion of an area previously belonging to the image background, while the false image the area determines background motion that the object has just left. Fig. 1 shows an example of this false motion detection.

In this example you may appreciate the motion of a bus through two images of a sequence, as well as the difference between two images of the sequence. Labelled under (A) we have the movement detected due to the object, while under (B) we may appreciate the false movement due to the image background.

Fig. 1. Object and background motion detection

When facing background motion, the lateral interaction mechanisms have to consider the influence of the environment of the motion detection areas, so that background motion may be diluted. A simple way to affront this task is adjusting the threshold values, in order not to allow accumulative computation or lateral interaction. This way it is possible to perfectly delimit valid charges in each pixel from valid or non-valid silhouette information.

3.5 The Motion Speed

When an object moves along an image, there are two possible cases (see Fig. 2). (1) The element moves at enough speed, or the sampling rate is slow enough, so that no pixel of the object at this instant occupies any pixel of the object in the previous instant. (2) This intersection is not null.

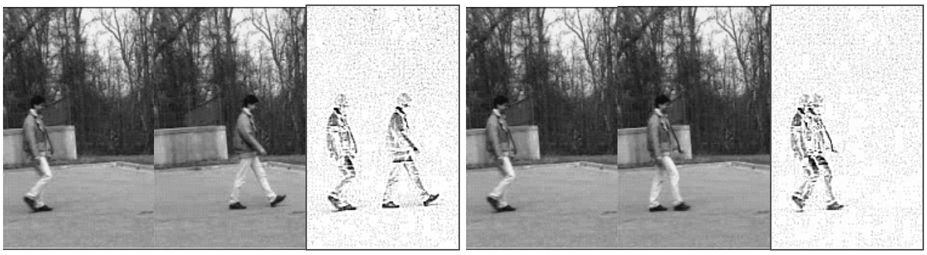


Image at t-Δt Image at t Image diff. Image at t-Δt Image at t Image diff.
(a) (b)

Fig. 2. Image difference (a) without and (b) with intersection

The case where the intersection is empty does not represent any problem, since the entire silhouette of the element is displayed (both in the previous instant *t-Δ t* and in the current instant *t*). This is not the case when this intersection is not empty.

With the purpose of considering this problem, the possibility to work with intermediate charge values ($v_{dis} < C(x,y,t) < v_{sat}$) has been incorporated. These intermediate charge values belong to those parts of moving objects that have not varied in their grey levels, but that we know they are part of the object.

By using an appropriate global computing function F_g, we can force the pattern (or part of the pattern) to start shaping while the moving element changes its position. Therefore, the solution to the problem of motion speed lays in incorporating an activation / deactivation mechanism that allows to centre the attention on where the element is moving, ignoring where the element comes from. And this is carried out by means of the election of a convenient global computing function F_g.

A necessary mechanism of lateral co-operation able to satisfy the following premises is implemented: (a) Pixels maximally charged ($C(x,y,t) = v_{sat}$) help to increase the charge of intermediate charged pixels ($v_{dis} < C(x,y,t) < v_{sat}$) that are directly or indirectly connected to them. (b) Pixels charged to intermediate values ($v_{dis} < C(x,y,t) < v_{sat}$) with no connection to any maximally charged pixels will tend to discharge, and therefore to total deactivation.

3.6 The Colour in Motion

Let us present in this section a simple and classic method of motion detection that allows knowing which image pixels of a sequence have experienced a variation in colour o grey level (being this latter the most usual). The algorithm used in numerous cases is:

$$MOV(x,y,t) = \begin{cases} 1, & if\ GL(x,y,t,) \neq GL(x,y,t-\Delta t) \\ 0, & otherwise \end{cases}$$

There is motion detection *MOV(x,y,t)* in an image pixel (*x,y*) at time instant *t*, if grey level *GL* has varied among instants *t-Δ t* and *t*. The method described is robust when the elements that move on the image sequence are composed of one single grey level. A binarisation results into a trivial way to differentiate among points that may or not have changed their grey level. The verified fact is that the changes in one grey level interfere with the changes in another one for the calculation of the element silhouettes.

It is easy to foresee that the motion detection method starting from the grey level change is valid when the object has a single grey level. So we have opted to segment each image in a number n (typically 8 or 16) of grey levels and to use the previous algorithms on each grey level band, as if we had *n* images formed by one single grey level moving objects. Notice that using all possible grey levels is equivalent to using 256 grey levels, what does not subtract generality to the exposed argument. Thus, there will be a charge value at each grey level band for each image pixel. In the same way, there a part of the silhouette at each grey level band and pixel will exist. At the end, shades obtained at each grey level band are gathered to obtain the complete silhouette of the moving object.

3.7 The Discrimination of More than One Moving Object

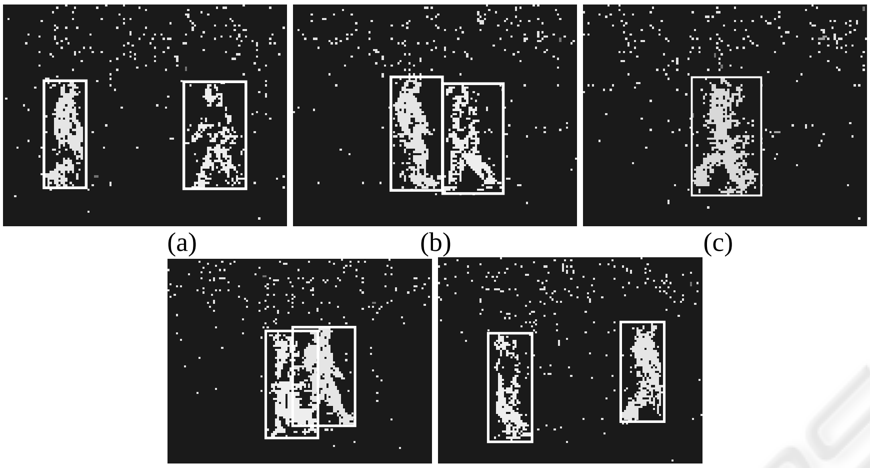
Another important problem related to motion in real image sequences is the high complexity that implies the calculation of all moving elements (objects) in the scene. In addition to this already serious problem is the discrimination or classification of the objects obtained. This is another non-trivial problem that is currently solved by silhouette searchers.

It is also proposed to use another lateral interaction mechanism to clearly differentiate among the silhouettes of all the objects in movement in the scene. It may be supposed that an object is a closed group of pixels with a charge value above a minimum required. Calculating the mean value of the charge of all the pixels that form the silhouette, a value is obtained which represents the charge in each instant *t*. This common charge value is the identity sign of the moving object. This value is evidently responsible for the spatio-temporal accumulative function F_l in local time scale *T*.

4 Data and Results

In order to show the functionality of the knowledge-based silhouette detection model described, next a significant example is offered. Here the image sequence *TwoWalk-New* downloaded from University of Maryland Institute for Advanced Computer Studies, copyright © 1998 University of Maryland, College Park, is used. This sequence was originally created to test the real time visual surveillance system W^4 [6]. It shows two people (a man and a woman) walking through a scene. Fig. 10 shows the result of applying the knowledge-based model proposed to some images of the sequence. The parameter values for this experiment were Δt=0.35 seconds, ΔT=0.005 seconds, n=8, v_{dis}=0, v_{sat}=255. Note that the model perfectly detects the silhouettes of the non-rigid objects in a quite simple manner.

Fig. 3 (a) offers the two silhouettes at a time instant close to the beginning of the sequence. The young man is walking from left to right, while the young woman is walking from left to right. Note that the man’s silhouette may be best appreciated. This is due to the greater contrast of the man’s cloth with the background.



(a) (b) (c) (d) (e)

Fig. 3. Some result images

There is also some noise present in the resulting images, especially in the upper part of the images. This is naturally due to the motion of the trees observed from the two last computed images.

Fig. 3 (b) presents the result before both people interact in the 2D projection of their 3D shapes. On Fig. 3 (c) you may see what happens during the occlusion phase. Both people appear as a unique silhouette. Remember, once again, that this is due to lateral interaction.

The situation on Fig. 3 (d) has now changed. The young people are no more on the same 2D projection. So there are two silhouettes again. This will be the situation through the rest of the sequence.

7 Conclusions

In this paper a model for lateral interaction in accumulative computation and its application to motion detection after introducing knowledge on motion detection has been proposed. The model incorporates accumulative computation, lateral interaction and double time scale, and can be considered as biologically plausible.

The model 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. In contrast to similar approaches no complex image preprocessing must be performed and no reference image must be offered to the model.

A background subtraction technique [7] could have been chosen as the general outline. But notice the uselessness of this technique in real environments of outdoor scenes. Methods based on the correlation of characteristics [8] have not been taken in account because they are excessively dependent on the type of application faced. Gradient-based methods [9], in turn, suffer from excessive problems to be able to pursue the needs of this work. The biggest inconvenience of these methods it is that they are excessively directed to the obtaining of one single pattern in the scene, and that they offer very poor results in the presence of diverse objects. Also, the computational cost is enormous. Neither the methods based on regions [10] fulfil the general requirements looked for in this work. They are methods excessively oriented to the search in very concrete regions of the images. It may also be highlighted that the proposed model has no limitation in the number of non-rigid objects silhouettes to differentiate. This knowledge-based model facilitates object classification by taking advantage of the object charge value, common to all pixels of a same moving element. Thanks to this fact, any higher-level operation will decrease in difficulty. The model seems to be promising in a lot of different applications related to image processing. The model is currently being tested in very different real world applications.

Acknowledgements

This work is supported in part by the Spanish CICYT TIN2004-07661-C02-02 grant.

References

1. E. Trucco, A. Verri, Introductory Techniques for 3-D Computer Vision, Prentice Hall, 1998.
2. A. Fernández-Caballero, Jose Mira, Ana E. Delgado, M.A. Fernández, “Lateral interaction in accumulative computation: A model for motion detection”, Neurocomputing, 50C, 341-364, 2003.
3. A. Fernández-Caballero, J. Mira, M.A. Fernández, A.E. Delgado, “On motion detection through a multi-layer neural network architecture”, Neural Networks, 16(2), 205-222, 2003.
4. J. Mira, A.E. Delgado, A. Manjarrés, S. Ros, J.R. Alvarez, “Cooperative processes at the symbolic level in cerebral dynamics: Reliability and fault tolerance”, in R. Moreno-Diaz and J. Mira, eds., Brain Processes, Theories and Models, The MIT Press, 244-255, 1996.
5. M.A. Fernandez, J. Mira, M.T. Lopez, J.R. Alvarez, A. Manjarrés, S. Barro, “Local accumulation of persistent activity at synaptic level: Application to motion analysis”, Springer-Verlag, LNCS, 930, 137-143, 1995.
6. I. Haritaoglu, D. Harwood, L.S. Davis, “W4: Who? When? Where? What? A real time system for detecting and tracking people”, Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 222-227, 1998.
7. S. Niyogi, E. Adelson, “Analyzing gait with spatiotemporal surfaces”, IEEE Workshop on Motion of Non-rigid and Articulated Objects, 64-69, 1994.
8. F.G. Meyer, P. Bouthemy, “Region-based tracking using affine motion models in long image sequences”, Computer Vision, Graphics and Image Processing: Image Understanding, 60(2), 119-140, 1994.
9. J.L. Barron, D.J. Fleet, S.S. Beauchemin, “Performance of optical flow techniques”, International Journal of Computer Vision, 12(1), 43-77, 1994.
10. M.J. Black, P. Anandan, “The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields”, Computer Vision and Image Understanding, 63(1), 75-104, 1996.