A Block-Based Human Model for Visual Surveillance

Encarnación Folgado, Mariano Rincón*, Margarita Bachiller, and Enrique J. Carmona

Departamento de Inteligencia Artificial. Escuela Técnica Superior de Ingeniería Informática, Universidad Nacional de Educación a Distancia, c/ Juan del Rosal 16, 28040 Madrid, Spain
mrincon@dia.uned.es

Abstract. This paper presents BB6-HM, a block-based human model for real-time monitoring of a large number of visual events and states related to human activity analysis, which can be used as components of a library to describe more complex activities in such important areas as surveillance. BB6-HM is inspired by the proportionality rules commonly used in Visual Arts, i.e., for dividing the human silhouette into six rectangles of the same height. The major advantage of this proposal is that analysis of the human can be easily broken down into parts, which allows us to introduce more specific domain knowledge and to reduce the computational load. It embraces both frontal and lateral views, is a fast and scale-invariant method and a large amount of task-focused information can be extracted from it.

1 Introduction

As well as for tracking, understanding human behavior is an important issue in intelligent visual surveillance. Behavior can be defined by activity composition[6,3]. To analyze and recognize these activities many human models have been proposed which have been divided into different groups depending on the representation used (see [4] for more details): stick models or skeleton models (2D, 3D and hybrid models)[1], geometric shape-based models (such as a simple box or elliptic shapes)[9], models based on significant points [7], deformable models [8], etc. The work by Haritaoglou et al. [5] is particularly highlighted, which uses different proportion aspects between the different body limbs. In this model the parts of the body and their position are statically specified.

Most of the human models found in the bibliography are primarily concerned with providing algorithms that resolve particular problems related to human modeling: detection of parts of the body, analysis of movement, detection of the pose, etc., and they are applied in many instances to surveillance tasks. Nevertheless, they do not treat the problem globally and they do not include real time aspects as constraints of their research. The solutions provided require

* Corresponding author.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 208–215, 2009.

a computational cost that is not very viable in most instances. Some of these models detect some parts of the body, but they do not treat the human as a whole, which implies that they may not be very effective in surveillance tasks. Others offer too simplistic treatment and merely inform of static characteristics. This work tries to cover the gap existing in these human models.

Our model is inspired by the proportionality rules commonly used in Visual Arts. It is a 2D model based on dividing the blob corresponding to the human silhouette, obtained at a stage of earlier segmentation [2], into areas determined by proportionality rules called “blocks’ (the division of the body into segments has been typically determined by visual characteristics like color). The major advantage of this proposal is that analysis of the human can be broken down into parts so that we can obtain information on different parts and “forget” about the rest. The advantages of our model are as follows: it treats movement from the frontal and lateral view, is a low computational-cost, scale-invariant method and a large amount of task-focused information can be extracted from it.

2 Block-Based Human Model Description

The model presented in this work consists of dividing the human blob vertically into 6 regions with the same height (see blocks B_1, ..., B_6 in Fig. 1). The blocks in this division correspond to areas related to the physical position of certain parts of the body when it does normal movements that we wish to detect in surveillance. Specifically, standing and in a position of repose, the correspondences are as follows: head is in B_1, shoulders are in B_2, elbows are in B_3, hands and hip are in B_4, knees are in B_5 and feet are in B_6. Besides, this division enables us to focus our attention on specific areas and ignore the rest. For example, we know that in the normal movement of the human, hands will be in blocks B_3 or B_4. This narrows the problem and reduces it to a local analysis of these blocks.

We distinguish in this model, as can be seen in Figure 1, two views: lateral and frontal. Intermediate views are treated according to the closer lateral or frontal view. In both instances the blocks are obtained in the same way. The distinction between both views is done by analyzing the blocks. Thus, for example, it is seen that changes in size of blocks B3 and B4 with the movement of the arms will be greater for the lateral case than the frontal case, or changes in size of blocks B_5 and B_6 with the movement of the legs will be greater for the lateral case.

Several parameters and significant points are used in the model: global parameters $H_T(t)$ and $W_T(t)$ are shown in Figure 2.a); block parameters $H_{Bi}(t)$, $W_{Bi}(t)$, $W_{Li}(t)$ and$W_{Ri}(t)$) are shown in Figure 2.b particularized for block B_4; and some secundary parameters $HC(t)$, $CW_i(t)$ and $S_i(t)$, which are defined below.

In each frame, we identify different significant points based on all the silhouette points belonging to each block. In the first place, the upper and lower points ($P_U(t)$ and $P_L(t)$) are defined, which delimitate the height of the set of blocks, $H_T(t)$, and enable us to establish the vertical division in the different blocks, $B_i(t)$, $i = 1..6$. All blocks have the same height ($H_{Bi}(t) = H_T(t)/6$). If y_i and

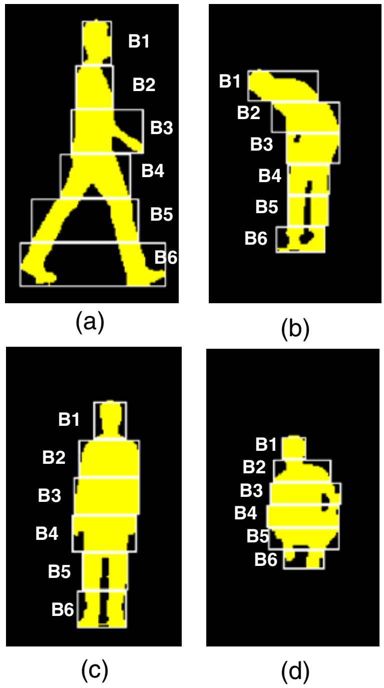

Fig. 1. A human's blob divided into blocks in lateral (a, b) and frontal (c, d) views

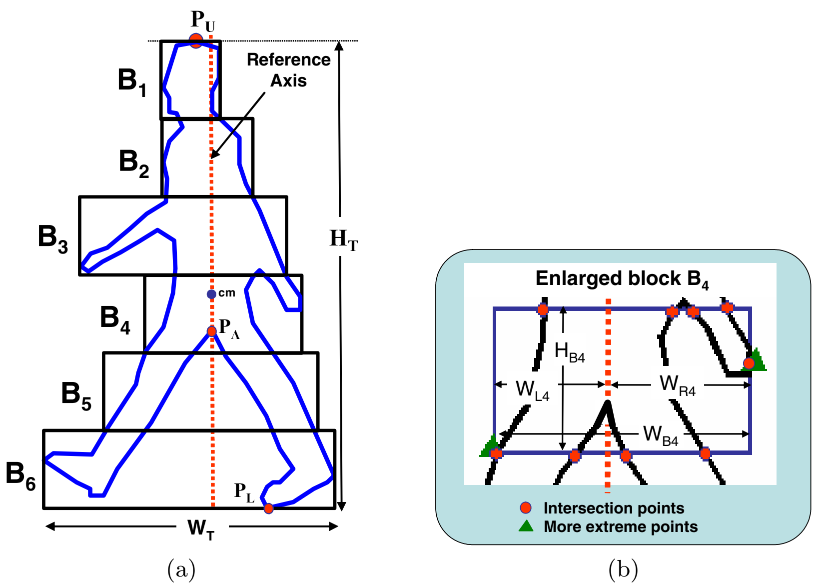

Fig. 2. Block-based human model parameters

y_{i+1} define the coordinates $(y - axis)$ of the upper and lower sides of the block $B_i(t)$, then the width of each block, $W_{Bi}(t)$, is delimited by the extreme left and right points of the silhouette fragment located between y_i and y_{i+1} . Also, for each block, the intersection points, which are the points belonging to the blob and which cut with some of the sides of $B_i(t)$, are obtained. Besides, a special significant point is the point joining the legs, $P_\Lambda(t)$. Finally, a reference axis is defined by the vertical line passing through the silhouette blob's center of mass (cm).

From this block model and assuming that the height of the human standing, H_S, is obtained from previous frames, a set of secondary parameters is defined related to different situations that we wish to detect:

- The height crutch relation (HC), which is a the relation between the height of the human upright, H_S, which is a static reference parameter, and the height of the point joining the two legs, $H_\Lambda(t)$. $H_\Lambda(t)$ is calculated as the vertical distance from $P_\Lambda(t)$ to B_6's lower side:

$$HC(t) = \frac{H_S}{H_S - H_\Lambda(t)} \tag{1}$$

  where $1 \leq HC < 2$.

- The change in width vector (CW), where each component contains the relation between the width of the block in a frame and the preceding one for each block B_i:

$$CW_i(t) = \frac{W_{Bi}(t)}{W_{Bi}(t-1)}, \qquad i = 1..6 \tag{2}$$

- The symmetry vector (S), where each component represents the proportion between the widths of the parts of the block B_i to the right and left of the reference axis:

$$S_i(t) = \frac{W_{Li}(t)}{W_{Ri}(t)}, \qquad i = 1..6 \tag{3}$$

The case model will consist of all the points and parameters characterizing the blocks: the significant points, global and block parameters, secondary parameters and reference parameters. The following section describes how to use this human model in different surveillance tasks.

3 Case Studies

The following subsections exemplify the information provided by BB6-HM.

Table 1. Parts of the human body located from the block model for a human standing upright in a lateral view

PART	DEFINITION
HANDS (P_{H1},P_{H2})	More extreme right and left points of blocks B_3 and B_4 (they are in B_4 in position of repose and in B_3 or B_4 when there is movement).
FEET (P_{F1},P_{F2})	More extreme right and left points of blocks B_5 or B_6 (they are in B_6 in position of repose and in B_5 or B_6 when there is movement).
HEAD (P_{HD})	Upper extreme point of block B_1 without the arms raised or midpoint of upper intersection points of block B_2 when arm or arms are raised over the head.
TORSO / BACK	Block situated immediately below the block to which the head belongs.

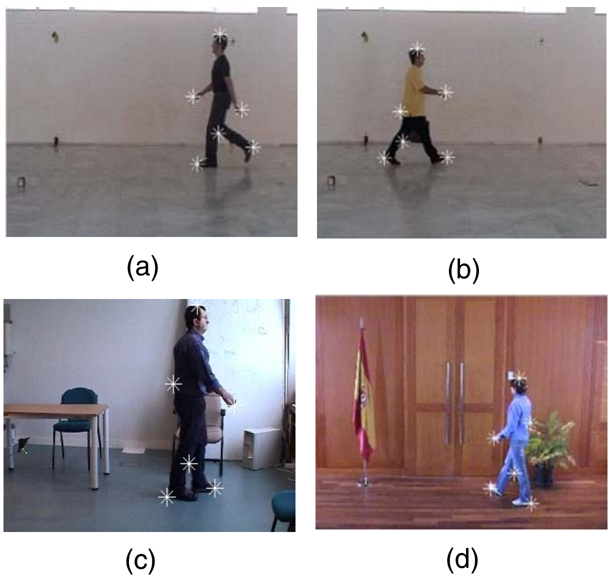

Fig. 3. Sample of sequences used for evaluation of the location of body parts in pure and partial lateral views of different humans carrying or not carrying objects (suitcase). The points P_{HD}, P_{H1}, P_{H2}, P_{F1}, P_{F2} and $P_{\Lambda}$are marked with " * ".

3.1 Location of Parts of the Body

One of the aims of the model described is to identify the position of the different parts of the human body. To locate them, the blocks are analyzed separately. Table 1 details a proposal for locating the main parts of the body and their association with the block model according to the position of repose or movement of the human standing upright in a lateral view.

Table 2. Percentage of correct locations of body parts in different video sequences

Seq. (No. of Frames)	P_{HD} (%)	P_{H1} (%)	P_{H2} (%)	P_{F1} (%)	P_{F2} (%)	$P_{\wedge}$ (%)
1 (90)	95.0	60.0	96.6	100	100	95.0
2 (60)	100	85.0	95.0	100	100	95.0
3 (36)	100	100	77.8	100	100	97.2
4 (73)	100	84.93	60.3	100	100	93.2
5 (73)	100	84.93	91.8	100	100	98.6
6 (100)	100	99.0	99.0	100	96.0	91.0
7 (47)	100	68.08	93.6	100	100	93.6
Total	99.1	82.7	88.9	100	99.2	94.5

This proposal has been evaluated on different types of sequences (see Fig. 3): man with (Fig. 3.b) and without (Fig. 3.a,c,d) suitcase, pure (Fig. 3.a,b) and partial (Fig. 3.c,d) lateral views, different scales and camera perspectives. Also in this figure, the points corresponding to the head, hands, feet and P_{Λ} are shown.

Table 2 shows the percentage of hits on the location of these body parts in the 7 sequences analyzed. As can be seen, accuracy in the location of the head is very high due to the fact that in no sequence did the humans raise their hands. In general, their feet are correctly found but the same is not true for their hands because they are occluded during walking. Finally, the location of point P_{Λ} is more sensitive to the morphological operations performed in the segmentation process. Even so, the results are quite satisfactory.

3.2 Recognition of Primitive States and Events

The model parameters can be used to define rules that will allow us to classify different events of interest [3]. For example, let us assume that we want to detect whether the person is carrying or not-carrying an object that he is holding in one

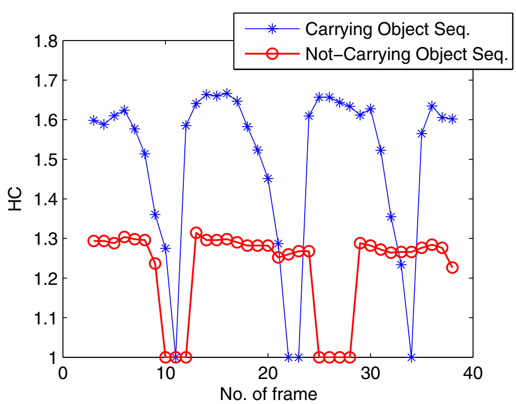

Fig. 4. Temporal evolution of the HC parameter: carrying and not carrying a suitcase

Table 3. CarryingObject event results

Seq. No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Carrying Object	no	yes	no	yes	yes	no	no
Max(HC)	1.60	1.36	1.67	1.39	1.39	1.67	1.48
CarryingObject event	0	1	0	1	1	0	0

of his hands. In a lateral view, this situation includes analyzing the parameter, HC. Figure 4 shows the temporal evolution of this parameter in two sequences, one with a human carrying a suitcase and another with no suitcase being carried.

The event CarryingObject is detected by the rule expressed in Eq. 4., where the threshold value $TH_{CO} = 1.45$ was heuristically selected. Table 3 shows the results on the sequences analyzed. The event is correctly detected in all the sequences. Note that this definition is valid only if the object is carried with one hand and the arm outstretched (suitcase):

$$CarryingObject = \begin{cases} 1 & if \quad max(HC) < TH_{CO} \\ 0 & otherwise \end{cases} \tag{4}$$

4 Conclusions

This work has presented BB6-HM, a new blob-based human model for visual surveillance. Besides being scale invariant and having low computational cost, the novelty of this model lies in breaking down the human silhouette into parts to impose a structure, which helps constrain the areas where significant points are located, thereby simplifying the analysis.

Examples of using this model on lateral views are shown. The modularity of the system facilitates the definition of new primitive states and events incrementally and simply.

This model can be used to build an event library related to human pose and movement to describe more complex activities that are of particular interest in surveillance tasks. To define these events, in some instances, analyzing static properties is enough, while in others, it will be necessary to do a dynamic analysis with a temporary window from the instant considered backwards in time. This dynamic analysis and recognition of human event and activity patterns with machine learning techniques are the issues currently being investigated by our research group.

Acknowledgments

The authors would like to thank the CiCYT for financial support via project TIN-20007-67586-C02-01 and the UNED project call 2006.

References

1. Ben-Arie, J., Wang, Z., Pandit, P., Rajaram, S.: Human Activity Recognition Using Multidimensional Indexing. IEEE Trans. on Pattern Analysis and machine Intelligence 24(8), 1091–1104 (2002)
2. Carmona, E.J., Martínez-Cantos, J., Mira, J.: A new video segmentation method of moving objects based on blob-level knowledge. Pattern Recognition Letters 29(3), 272–285 (2008)
3. Carmona, E.J., Rincón, M., Bachiller, M., Martínez-Cantos, J., Martinez-Tomas, R., Mira, J.: On the effect of feedback in multilevel representation spaces for visual surveillance tasks. Neurocomputing 72(4-6), 916–927 (2009)
4. Gavrila, D.: The Visual Analysis of Human Movement: A Survey. Computer Vision and Image Understanding 73(1), 82–98 (1999)
5. Haritaoglu, I., Harwood, D., Davis, L.S.: W4: Real-Time Surveillance of People and Their Activities. IEEE Trans. on Pattern Analysis and machine Intelligence 22(8), 809–830 (2000)
6. Martinez-Tomas, R., Rincón, M., Bachiller, M., Mira, J.: On the Correspondence between Objects and Events for the Diagnosis of Situations in Visual Surveillance Tasks. Pattern Recognition Letters 29(8), 1117–1135 (2008)
7. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfinder: Real-Time Tracking of the Human Body. IEEE Trans. on Pattern Analysis and machine Intelligence 19(7), 780–785 (1997)
8. Zhang, J., Collins, R., Liu, Y.: Representation and Matching of Articulated Shapes. In: International Conference on Computer Vision and Pattern Recognition, pp. 342–349 (2004)
9. Zhao, T., Nevatia, R.: Tracking Multiple Humans in Complex Situations. IEEE Trans. on Pattern Analysis and machine Intelligence 26(9), 1208–1221 (2004)