A block-based model for monitoring of human activity

Encarnación Folgado, Mariano Rincón*[1], Enrique J. Carmona, Margarita Bachiller

Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia,
C/ Juan del Rosal 16, 28040, Madrid, Spain.

Abstract— The study of human activity is applicable to a large number of science and technology fields, such as surveillance, biomechanics or sports applications. This article 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, for example, luggage at airports, clients' behaviour in banks and patients in hospitals. 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 regions, so that we can obtain information of activities. The computational load is very low, so it is possible to define a very fast implementation. Finally, this model has been applied to build classifiers for the detection of primitive events and visual attributes using heuristic rules and machine learning techniques.

Index Terms: Human models, Human activity recognition, video-based surveillance, Video analysis, primitive visual events.

1 INTRODUCTION

Nowadays it is very usual for public and private companies to have sophisticated surveillance systems that try to monitor the state of their "business" in order to detect anomalous situations and avoid adverse or undesirable situations. The final aim of an artificial vision system is to provide a task-focused scene description. Concentrating on the description of human activity, and in accordance with [25][23][14], the recognition of activities is considered as a problem of classifying spatial-temporal characteristics or, alternatively, as "queries" on higher-level constructions. In other words, activities are considered as complex events, which are defined by spatial-temporal composition of simpler events, and these simple events in turn are defined by other even simpler events and so on to form a hierarchy of events to link with primitive events, which are determined from changes in state of visual attributes.

This article presents BB6-HM, a model that can be used to monitor, in real time and with a minimum computational load, a large number of these primitive events related to humans, which can be used as components of a library to describe more complex activities in many surveillance tasks (luggage at airports, clients' behaviour in banks, patients in hospitals, etc.). Finally, this model has been applied to build classifiers for the detection of primitive events and visual attributes using heuristic rules and machine learning techniques.

* Corresponding author: e-mail address: mrincon@dia.uned.es

This article is organised as follows. The second section analyses other works related to the proposal. The third section describes the block-based human model. The fourth section describes some events and visual attributes detected using relations between the parameters of the human model and evaluates on some test sequences. Finally, the fifth section presents the conclusions and highlights future works.

2 RELATED WORK

This section gives some works related to models of humans which analyse the type of movement done. Some of them, like the model in this work, focus on surveillance. They have been divided into different groups depending on the model used.

The first group includes those works represented by the "stick models" or skeleton models. In these models the human body is represented as a set of segments (bars or volumes) that are joined together with articulations. This representation is based on the observation that human movement is essentially the skeleton's movement aided by the adjacent muscles. The geometry of each of the segments varies according to authors and depends on the application (for example: 2D or 3D tracking). Thus, as 3D models we can mention [4][10][9] and as 2D models [13],[32],[28],[2]. There are also some hybrid applications, which mix both representations [27][3]. Other work is this by Fujiyoshi and Lipton [13], which method consists of generating what is called star skeleton of the silhouette, tracing this star from the centre of mass to the extremities (feet and hands) and the head. Others are based on looking for the axes of a silhouette and extracting the skeleton by tracing a line between parallel axes, see Niyogi and Adelson [26]. Some research aims to detect a dynamic situation, such as walking, running, etc. One way of detecting this situation is to use transformations on the middle axis [1] or using transformations of the distance like [20]. This last model is more advanced and allows those parts that are not very interesting to be removed and focuses on those that provide the information that is required. For example, if it is necessary to look for the centre of the human masses, the arms and legs are not interesting. Gavrila [15] uses the skeleton models, locating the torso first, and using it to restrict the search of the extremities by breaking down the space of states. The bar model by Chen and Lee [7] contains 17 segments and 14 articulations that represent the characteristics of the head, torso, hip, arms and legs. Deutscher [12] proposes a model of 17 segments with 29 degrees of freedom, where each segment is a cone of elliptic section.

All these human models have disadvantages. On the one hand, the excessive simplicity of some mean that they are not very operative for surveillance and are basically used for determining the human pose. On the other hand, there are more versatile models, especially those that work in 3D, which can be applied to the surveillance field. These have the advantage that they manage occlusions very well and obtain more significant data on the analysis of the scene. Nevertheless, among their disadvantages is the difficulty of preparing these quite sophisticated models and their high computational cost. This makes them inadvisable in surveillance tasks that work in real time or need to compute a large amount of data.

Another category of models is where the modelling of the human is done with geometric shapes. A human can be represented with a simple box, as is described in Darrell [11]. This representation is basically used as an intermediate representation and not as a definitive representation. Instead of this, the elliptic shapes are used much more [24][29][33]. These types of representations can be applied in many tracking tasks, mainly in surveillance and they make it possible to evaluate the orientation of the human when several cameras are used.

Another group is formed by models based on representing significant points. Within this group, a well-known way of modelling consists of representing the human with three points: one for the head, and the other two for each hand. This way of representing the human is quite compact and makes it possible to describe most of a human's usual movements. Wren et al. [31] use this representation in the *Pfinder* system, which is a real time system to track people and interpret their behaviour. It is also currently used in different surveillance systems. Another compact representation is presented in [21], where the omega-shape features of people's head-shoulder parts is used for detecting and tracking humans. In another work [8], humans are represented with more points: head, shoulders, hands, armpits and feet. This model is applied to surveillance systems and focuses on specific parts of the human body (face, hands, ...etc.) which require a high computational load.

A set of hybrid works is presented below, which combine the aforementioned representations. The work by Haritaoglou et al. [16][17][18] is also highlighted, which uses different proportion aspects between the different body limbs. The proportion is used between the head, hands, torso and feet obtaining six fundamental parts and ten secondary parts to find a representation in the form of regions. The model specifies the parts of the body and their position statically.

Most of the works mentioned above are mainly concerned with providing algorithms that resolve the problems related to human modelling: detection of parts of the body, analysis of movement, detection of the pose, etc., applied in many instances to surveillance tasks. Nevertheless, they do not treat the problem globally and they do not include performance aspects as aims of their research. The solutions provided require, in many instances, a very high computational cost. Many of these works 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, like the kind of movement or pose, which does not provide enough information for a surveillance system. Others, on the other hand, are too complex to work in 3D, which make the computational cost very high and they are very slow and not very useful in applications that require real time, like surveillance.

Our proposal is in line with [22], where a system based on simple rules is described for recognition of human posture in home care scenarios. They also assume that the human silhouette has been obtained previously and that the human posture can be classified from some parameters and simple rules. Another interesting reference is [19], where automatic learning is used to define a system for gender classification from multiple views.

3. MODEL DESCRIPTION

The model presented in this work consists of vertically dividing the human blob into 6 regions with the same height (see blocks B_1, ..., B_6 in Fig. 1). Each of these regions can be delimited by a bounding box which we will call "block". This model is inspired by the proportionality rules commonly used in Visual Arts. In this case, the blocks of this division correspond to areas related to the physical position of certain parts of the body. For example, 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. This division enables us to focus our attention on specific areas and ignore the rest. For example, if we are analysing hands, we know that in a normal situation these are between blocks B_3 or B_4. This limits the problem and reduces it to a local analysis of these blocks.

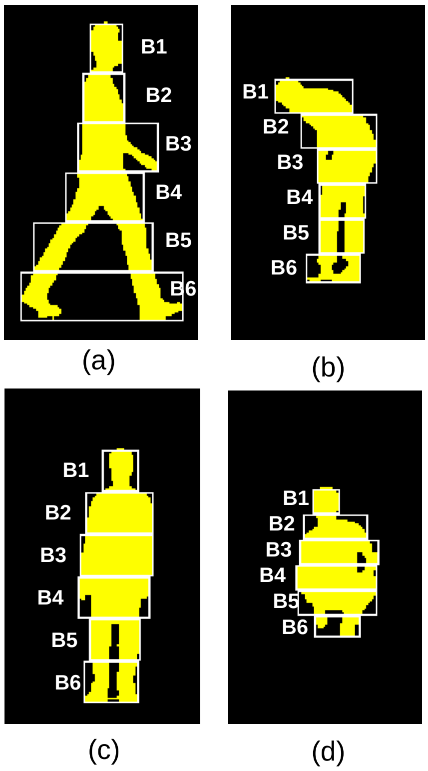
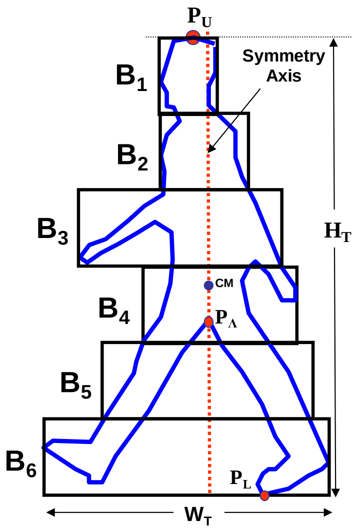


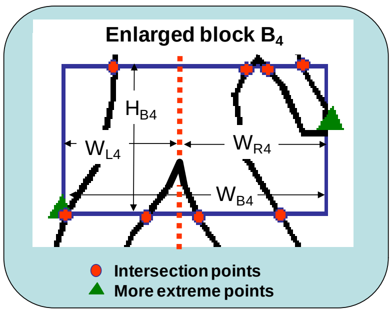
Fig. 1. A human blob divided into blocks in lateral (a, b) and frontal (c, d) views.

This model deals with *lateral* and *frontal* views, and it is not intended to analyse *zenital* view. In a frontal view the human is looking to the camera or has his back to it. In a lateral view, the human is laterally to the camera. These views can be measured by the view angle: a frontal view around 0º and 180º and a lateral view for the rest of angles. In both instances the model parameters are obtained in the same way.

We will use Fig. 2 to help us define the parameters and significant points used in the model. General parameters are shown in Fig. 2.a, while parameters defined for each block are shown in Fig. 2.b particularized for block B_4. We identify different significant points based on the silhouette points belonging to each block. In the first place, the upper and lower points (P_U and P_L) are defined, which limit the height of the set of blocks, H_T, and enable us to divide it in the different blocks, B_i, i=1...6. All blocks have the same height ($H_{Bi}= H_T/6$). If y_i and y_{i+1} define the y-coordinates of the upper and lower sides of the block B_i, then the left and right sides of each block are delimited by the extreme left and right points of the silhouette fragment located in block B_i, between y_i and y_{i+1}. Now we can define a new parameter: the width of each block, W_{Bi}.



(a)



(b)

Fig. 2. S*ignificant* points and primary parameters of the *block-based human* model (BB6-HM): a*) general view; b)* Enlarged view of block B4.

A symmetry axis is defined by the vertical line that passes through the centre of mass of the entire blob (*CM*). This axis allows us to measure the block width on each side (W_{Li},W_{Ri}). In order to normalize the width parameters with respect to the distance to the camera, we divide parameters W_i, W_{Li} and W_{Ri} by H_T.

For each block, the intersection points, which are the points belonging to the silhouette and which cut with some of the sides of B_i, are obtained. Finally, a special significant point is P_{Λ}, the junction point of the silhouette of the legs in its inner side (see Fig. 2.a).

From this initial points and measures, a set of secondary parameters are defined related to different situations that we wish to detect:

- The *change_in_width* vector (CW^T)

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

where each component contains the relation between the block width in a frame t and the preceding t-T frame for each block B_i.

- The *change_in_mass_centre* vector (ΔCM^T):

$$\Delta CM^T = (CM_x(t) - CM_x(t-T), CM_y(t) - CM_y(t-T)) \qquad (2)$$

where t represents the current frame instant, t-T represents the preceding T frame and CM_x and CM_y are the x and y coordinates of CM, respectively.

- The *directional symmetry* vector (S):

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

where each component represents the proportion between the widths of the parts of the block B_i to the right and left of the symmetry axis.

- The *symmetry* vector (S):

$$S_i = \frac{min(W_{Li}, W_{Ri})}{max(W_{Li}, W_{Ri})}, i = 1..6 \qquad (4)$$

where each component represents the adirectional symmetry proportion between the widths of the parts of the block B_i to the right and left of the symmetry axis.

- *The height_crutch* relation (HC):

$$HC = \frac{H_T}{H_T - H_\wedge} \qquad (5)$$

where H_Λ is calculated as the vertical distance from P_Λ to lower side of B_6.

- *Swinging feet* coefficient (S_f):

$$S_f = \frac{max(W_{L5}, W_{L6},) + max(W_{R5}, W_{R6})}{H_T} \qquad (6)$$

This parameter is used for detecting activities related to gait. If we assume that feet are the extreme points of blocks B_6 or B_5, this coefficient is a measure of the legs aperture.

- *Swinging-hands* coefficient (S_h):

$$S_h = \frac{max(W_{L3}, W_{L4},) + max(W_{R3}, W_{R4})}{H_T} \tag{7}$$

This is an important parameter when studying periodicity, type of movement, and different actions associated with the movement of the arms.

4. MONITORING OF HUMAN ACTIVITY

In order to use the BB6-HM model, the system shown in Fig. 3 has been used. With this system it is possible to detect primitive states and events from a video sequence where the human is segmented. In an initial stage, a description of the human is obtained (*human description)* containing constant and variable parameters. Variable parameters form the *case model* and characterise the human in a specific instant, t. These parameters will be obtained from the human's blocks characteristics. The constant parameters (*reference model*) define characteristics of the human in a known situation and act as a reference to characterise the temporary situation of the human more abstractly. The events of interest ("*Primitive Event Descriptions*") are characterised with spatial-temporal relations on the human's model parameters. Therefore the detection of the events consists of simply recognising each of their associated patterns, which makes it very useful since a modular and extensible system is obtained.

The following subsections detail how to use the information provided by the block-based human model described in section 3. This information can be used to build a library of primitive states and events and which can be used to describe more complex activities done by humans and of particular interest in surveillance tasks. Concretely we have focused on the location of parts of the body, the determination of the view of the human with respect to the camera and the detection of humans carrying objects.

Each of the situations recognised is described as a function of the model parameters (primary and secondary parameters). In some instances, just analysing the spatial properties of the silhouette in the time instant, t, is enough, while in others, it is necessary to do a spatial-temporal analysis. In this paper we use static relationships among parameters, but it is clear that more situations could be characterized with a dynamic analysis of the scene.

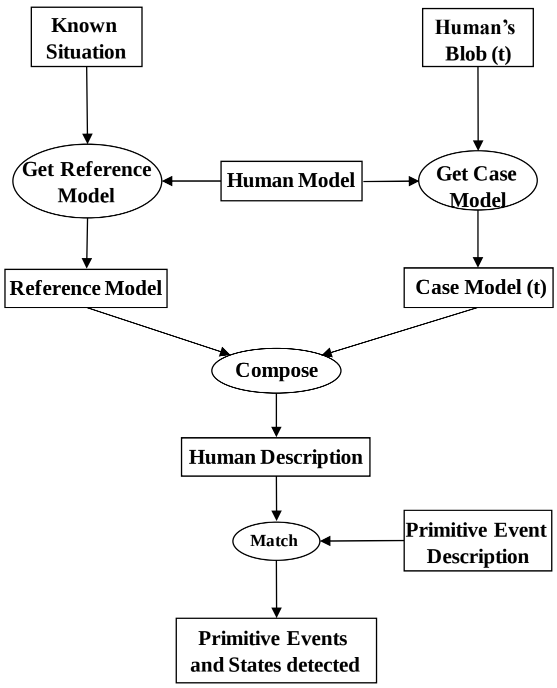


Fig. 3. Diagram of the system for recognizing primitive events from video sequences with the human model.

To evaluate the usefulness of the system, two kinds of video data have been used. On the one hand, the parts of the body have been located on several sequences, where we have used our own background subtraction method [6] to obtain the human blob. A sample of these sequences is shown in Fig. 4: man with (Fig. 4.b) and without (Fig. 4.a,c,d) suitcase, pure (Fig. 4.a,b) and partial (Fig. 4.c,d) lateral views and different scales and camera perspectives. On the other hand, the determination of the view of the human with respect to the camera and the detection of humans carrying objects have been carried out on Casia data base [5] using decision trees. Specifically, we have used the J48 algorithm implementation of Weka [30]. The Casia database sequences include views in angle increments of 18 degrees (0º, 18º, 36º, ..., 180º) and that distinguishes three classes related with the movement of humans carrying or not objects: class 1) human is wearing normal clothing (tight to the body) and doesn't carry objects; class 2) human is wearing a coat and doesn't carry objects and class 3) human is wearing normal clothing and is carrying an object (briefcase or bag). Due to the huge size of the database (124 individuals), 11 randomly selected individuals (124,000 instances) were considered for learning the classifier.

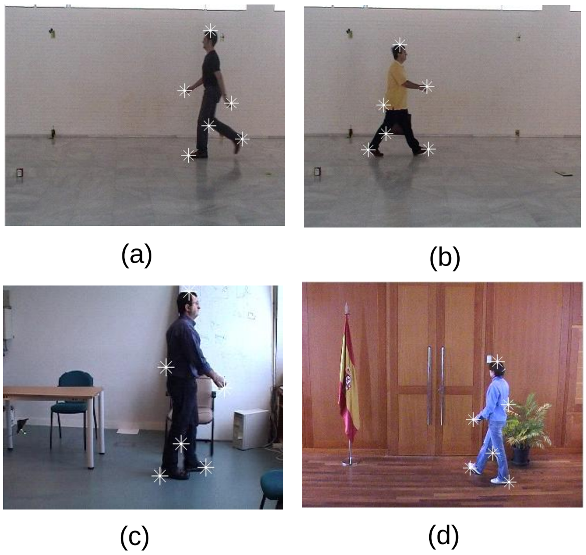


Fig. 4. Sequence examples used for locating parts of the body. The points P_{HD}, P_{H1}, P_{H2}, P_{F1}, P_{F2} and $P_{\wedge}$ are marked with "*".

4.1 LOCATION OF PARTS OF THE BODY

One of the aims of the system described is to identify the position of the different parts of the human body. To locate them, the blocks are analysed separately, depending on the part of the body that interests us. Starting from the proportionality of the human body and the known pose of the human in rest situation, the correspondence existing between each of the blocks and the different parts of the body is known. In the case of movement, the adjacent blocks to those analysed in the rest situation will also have to be analysed. Table 1 details the heuristic rules used for associating the parts of the body with the specific model blocks according to the situation of rest or movement of the human. These definitions are valid for frontal and lateral.

PART	DEFINITION
HANDS (P_{H1}, P_{H2})	Extreme right and left points of blocks B_3 or B_4: in the position of repose in B_4 and when there is movement in B_3 or B_4.
FEET (P_{F1}, P_{F2})	Extreme right and left points of blocks B_5 or B_6: in the position of repose in B_6 and when there is movement in B_5 or B_6.
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.

Table 1. Parts of the human body located from the BB6-HM for a human standing upright.

In Fig. 4, 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 parts of the body in seven sequences. 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 sometimes occluded during walking. Finally, the location of point $P_{\wedge}$ is more sensitive to the morphological operations performed in the segmentation process. Even so, the results are quite satisfactory. The average success percentage varies between 82.7% and 100%.

Sequence (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	84.9	60.3	100	100	93.2
5 (73)	100	84.9	91.8	100	100	98.6
6 (100)	100	99.0	99.0	100	96.0	91.0
7 (47)	100	68.1	93.6	100	100	93.6
Average	99.1	82.7	88.9	100	99.2	94.5

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

4.2 FRONTAL AND LATERAL VIEWS AND VIEW ANGLE

Two different experiments to assess the orientation of the human with respect to the camera have been carried out. On the one hand, we have generated a model to distinguish the frontal and lateral views, because the meaning of the model parameters for the detection of various events using heuristics strongly depends on it. On the other hand, another model to determine the exact angle formed by the individual with respect to the camera has been generated.

Frontal and lateral view classification

For training the J48 classifier, the Casia database has been used, making the correspondence between 0 and 180 degrees with *frontal view* class and the rest with a *lateral view* class. From the model parameters described in Section 3, various experiments have been conducted to determine the ones most discriminating. The best results were obtained from the widths of the blocks B_2, B_5 *and* B_6 (W_2, W_5, W_6), obtaining an accuracy of 91.9% over the test set. As we can see, the information is contained in a small number of parameters. Moreover, the error occurs in images with a bad segmentation, which makes incorrect the information associated with some of the parameters of the model.

View angle classification

From the model parameters described in Section 3, various experiments have been conducted using the J48 classifier to determine the most discriminating parameters. The best results were obtained from the following feature vector: (W_2, W_5, W_6, DS_2, DS_5, DS_6, ΔCM^6_x, ΔCM^6_y), obtaining a classification accuracy of 78.0% over the test set.

In Table 3 the confusion matrix is shown. The main diagonal indicates the number of correctly classified instances achieved on the test set. As for the errors, we can see that most of the error is clustered in the neighbour angles (the first diagonals below and above the main diagonal) and a small part of the error corresponds to the supplementary angle (data organized around the secondary diagonal). In some applications it is sufficient to indicate a rough orientation, therefore, if we consider as correct the angle with a deviation of ±18º, the success rate (sum of main diagonal and adjacent diagonals) is 90.5%. Moreover, as in the previous case, we must take into account the segmentation errors.

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	<-- classified as
3509	87	21	0	0	0	0	0	3	6	237	a = 0
88	3012	344	59	1	0	1	20	74	175	65	b = 18
24	362	2757	326	36	13	12	44	80	144	17	c = 36
0	47	374	2890	158	69	86	124	100	36	0	d = 54
0	2	34	147	2905	427	225	114	49	8	0	e = 72
0	2	21	73	484	2861	358	62	12	0	0	f = 90
0	1	26	75	255	405	2824	262	62	2	0	g = 108
0	14	43	129	117	85	247	2751	335	36	0	h = 126
3	53	59	85	39	25	91	376	2966	251	5	i = 144
6	159	132	31	14	1	6	55	233	3179	26	j = 162
274	38	13	0	0	0	0	0	9	24	3461	k = 180

Table 3. Confusion matrix for the view angle classifier

4.3 CARRYING OBJECTS

In this last experiment, the ability to detect whether the human carries objects has been evaluated. Initially, the parameter HC was used because it seemed to be the most significant parameter for determining this state. In Fig. 5, we analyse the influence of this parameter for the detection of the event carrying-object. In this figure, the temporal evolution of the HC parameter in a lateral view when the human carries a briefcase and when not is shown. Note how the temporal evolution is periodic in both cases, but the fundamental difference is in its amplitude, which turns out to be much greater when the human is not carrying object. However, the problem solution can not be reduced to the study of this parameter, because it is not valid in frontal view or when the human is carrying another type of object. Therefore, machine learning (J48 algorithm) has been used to define a model to classify such an event.

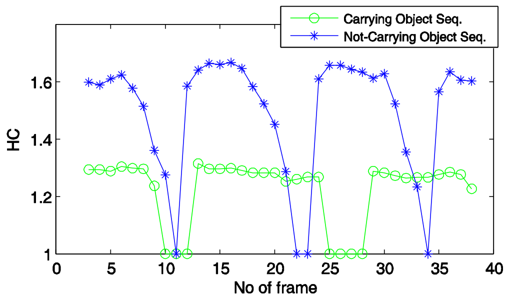


Fig. 5. Temporal evolution of the *HC* parameter in lateral view: carrying and not carrying a suitcase.

As our objective is to recognize whether or not a subject carries an object, we have defined two classes based on the Casia database: *no-carrying-object* (class 1 and 2 of Casia) and *carrying-object* (class 3 of Casia). From the model parameters, various experiments have been conducted to determine the most discriminating parameters. The best results were obtained from the following feature vector: (HC, W_2, W_3, W_4, W_5, W_6, S_2, S_3, S_4, S_5, S_6, S_f, S_h), obtaining a classification accuracy of 86.0% over the test set. As we can see, the *HC* parameter continues being a discriminating parameter. Note than, in this case, the non-directional symmetry parameter is more discriminative than the directional symmetry one.

5. CONCLUSIONS AND FUTURE WORKS

This work has presented a new human model (BB6-HM) oriented to surveillance. It embraces both frontal and lateral movements with a low computational cost and a large amount of task-focused information can be extracted from it. Furthermore, the orientation from the start of the system to the surveillance task allows us to indicate situations particularly useful for this task.

The experimental results have demonstrated the usefulness of the human model for the detection of primitive events and visual attributes. In particular, heuristic rules have been defined for the detection of parts of the body with satisfactory results, although dependent on a good segmentation. The BB6-HM parameters also allowed us to determine, through decision trees using a feature vector of only three components, frontal and lateral views with a classification accuracy of 91.9%. Even, the view angle was evaluated with a resolution of 18 degrees, obtaining a success rate of 78%. Finally, we built a new classifier applied to the detection of the *carrying-object* event, obtaining a success rate of 86%.

In future work, the model will be applied to the detection of primitive events and attributes visuals that allow monitoring of human activity using various pattern recognition techniques both static and dynamic. The information provided by the BB6-HM from different points of view could be integrated in order to obtain a 3D description of the human and thus increase the reliability of the system and recognise a greater number of events.

ACKNOWLEDGMENT

Support for this research was made possible from CICYT through TIN2007-67586-C02-01 project and UNED project call 2006 titled "Development of a knowledge based tracking system", in the context of which this study has been carried out. Portions of the research in this paper use the CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences.

6. REFERENCES

[1] A.G Bharatkumar, K.E Daigle, M.G Pandy, Q. Cai, and J.K Aggarwal. Lower Limb Kinematics of Human Walking with the Medial Axis Transformation. *In Workshop on Motion of Non-Rigid and Articulated Objects,* Austin, Texas, USA (1994)

[2] A.F. Bobick and J.W. Davis. The Recognition of Human Movement Using Temporal Templates. *IEEE Trans on Pattern Analysis and Machine Intelligence, vol.23, no. 3, pp.257-267*. (2001)

[3] B. Boulay, F.Bremond and M. Thonnat. Human Posture Recognition in Video Sequence. *VS-PETS, pp. 23-29*. (2003)

[4] B. Boulay, F.Bremond and M. Thonnat. Posture Recognition with a 3D Human Model. *ICDP 2005, pp. 135-138.* (2005)

[5] CASIA Gait Database. http://www.sinobiometrics.com

[6] E.J. Carmona, J. Martínez-Cantos and J. Mira. A New Video Segmentation Method of Mobile Objects Based on Blob-Level Knowledge. Pattern Recognition Letters, Vol. 29(3), pp. 272-285. (2008)

[7] Chen, Z; Lee and H. J. Knowlegde-guided visual perception of 3D human gait from a single image sequence. *IEEE transactions on Systems, Man and Cybernetics, 22(2) pp 336-342.* (1992)

[8] J.M. Chung and N. Ohnishi. Cue Circles: Image Feature for Measuring aumberg 3-D Motion of Articulated Objects Using Sequential Image Pair. *In Active ternational Conference on Automatic Face- and Gesture-Recognition*, Japan, (1998)

[9] P. Chunhong and M. Songde 3D Motion Estimation of Human by Genetic Algorithm. *15th International Conference on Pattern Recognition (ICPR'00) - Volume 1* p. 1159. (2000)

[10] I. Cohen and H. Li. Inference of Human Postures by Classification of 3D Human Body Shape. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 74-81*.(2003)

[11] T. Darrell, P. Maes, B. Blumberg, and A.P. Pentland. A Novel Environment for Situated Vision and Behavior. *In Workshop for Visual on Behaviors at CVPR-94,* (1994)

[12] J. Deutscher and R. Blake. Articulated body motion capture by annealed particle filtering. *Computer Vision and Pattern Recognition, Proceedings. IEE Conference on, Vo. 2 pp:126-133.* (2000)

[13] H. Fujiyoshi and A.J Lipton. Real-Time Human Motion Analysis by Image Skeletonization. *In Workshop on Applications of Computer Vision*,(1998)

[14] D. Gavrila. The Visual Analysis of Human Movement: A Survey. Computer Vision and Image Understanding, Vol. 73, No. 1, pp. 82-- 98. (1999)

[15] Gavrila, D, Davis, L. 3D model-based tracking of human in action: a multiview approach. *Proc Conf. Computer Vision and Pattern Recognition, 73-80.* (1996)

[16] I. Haritaoglu, D. Harwood and L.S. Davis. Ghost: A Human Body utter Part Labelling System Using Silhouettes. *In International Conference Recognition, on Pattern Recognition,* (1998)

[17] I. Haritaoglu, D. Harwood and L.S. Davis. W4: Who? When? Where? What? - A Real Time System for Detecting and Tracking - People. *In International Conference on Automatic Face- and Gesture Recognition, Nara,* Japón, (1998)

[18] I. Haritaoglu, D. Harwood and L. S. Davis. W4: Real-Time Surveillance of People and Their Activities. IEEE Trans. On Pattern Analysis and machine Intelligence, Vol. 22, No. 8. (2000)

[19] G. Huang and Y. Wang. Gender Classification Based on Fusion of Multi-view Gait Sequences. Y. Yagi et al. (Eds.): ACCV 2007, Part I, LNCS 4843, pp. 462–471. (2007)

[20] S. Iwasawa, K. Ebihara, J. Ohya, and S. Morishima. Real-Time Estimation of Human Body Posture from Monocular Thermal Images. *In Conferenceon Computer Vision and PaternRecognition*,(1997)

[21] M. Li, Z. Zhang, K. Huang and T. Tan. Rapid and robust human detection and tracking based on Omega-shape features. ICIP-2009 - IEEE International Conference on Image Processing. (2009)

[22] C. Li and Y. Chen. Human Posture Recognition by Simple Rules. SMC'06-IEEE International Conference on Systems, Man and Cybernetics Vol. 4, pp. 3237 – 3240. (2006).

[23] R. Martinez-Tomas, M. Rincón, M. Bachiller, J. Mira. On the Correspondence between Objects and Events for the Diagnosis of Situations in Visual Surveillance Tasks. Pattern Recognition Letters. Elsevier . Vol. 29, No. 8. pp. 1117 -1135. (2008).

[24] A. Nakazawa, H. Kato, and S. Inokuchi. Human Tracking Using Distributed vision systems. In International Conference on Pattern Recognition, (1998)

[25] R. Nevatia, S. Hongeng, and F. Bremond. Video-based event recognition: activity representation and probabilistic recognition methods. Computer Vision and Image Understanding, Vol. 96, No. 2, pp. 129-162, (2004)

[26] S.A. Niyogi and E. H. Adelson. Analyzing and Recognizing Walking Figures in XYT. *In Conference on Computer Vision and Pattern Recognition*. (1994).

[27] E. Ong and S. Gong. A Dynamic Human Model using Hybrid 2D-3D Representations in Hierarchical PCA Space. *Westfield College London* (1999).

[28] D. Ramanan, D.A. Forsyth and A. Zisserman. Tracking People by Learning Their Appearance. Trans on Pattern Analysis and Machine Intelligence, Vol. 29, No. 1, pp.65-81. (2007).

[29] A. Utsumi, H. Mori, J. Ohya, and M. Yachida. Multiple-View-Based gid Tracking of Multiple Humans. *In International Conference on Pattern Recognition*, (1998)

[30] WEKA Project: http://www.cs.waikato.ac.nz/~ml/index.html

[31] C.R. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland. Pfinder: SA, Real-Time Tracking of the Human Body. *In International Conference on Automatic Face- and Gesture-Recognition, Killington, Ver- er mont, USA*, (1996).

[32] J. Zhang, R. Collins and Y. Liu: Representation and Matching of Articulated Shapes, *CVPR* (2004)

[33] T. Zhao and R. Nevatia, , "Tracking Multiple Humans in Complex Situations", PAMI, Vol. 26, No. 9, pp. 1208-1221, (2004)