Modelling Postures of Human Movements

Djamila Medjahed Gamaz, Houssem Eddine Gueziri, and Nazim Haouchine

Computer Science Departement, USTHB University,
BP 32 El Alia, 16 000, Algiers Algeria
Tel.: +213 21 24 79 50 to 60
dmedjahed@usthb.dz, dmedjahedgamaz@yahoo.com

Abstract. The goal of this paper is to present a novel modelling of postures of human activities such us walk, run… Effectively, human action is, in general, characterized by a sequence of specific body postures. So, from an incoming sequence video, we determine the postures (key-frames) which will represent the movement. We construct the prototypes corresponding to these key-frames by thinning these postures, and then we use this skeleton as a starting point for building the model. Some results are presented to validate our models.

Keywords: Human Activities, Modelling, Shape Matching, Skeleton, thinning.

1 Introduction

Lot of papers present an overview of human motion estimation and recognition [1] [2] [3]. The video of measuring shape deformation relative to prototypes has a long history in pattern recognition and computer vision [4] [5]. The work in [6] [7] present an algorithm for computing correspondence between arbitrary shapes. Based on skeletons directly, many approaches have been developed for shape matching [8] [9] [10]. The benefits of applying skeleton-based methods are its natural consistency with human intuition and capability to describe the local geometrical features, allowing the performance of articulated matching [11] [15] [16]. In this paper, we present a novel modelling of human activities postures.

Inspired by works of [12] [13] [14] [18], we propose to hide (superimposed) the skeleton (of different body poses) on models representing human activities in a predefined database, to recognize the motion in the input video made by a single person. So, we first, construct prototypes of postures which describe a movement from an incoming sequence video. We determine automatically the postures which will represent the movement, we skeletal these postures then we use this skeleton as a starting point for building the model.

We test the relevance of the models constructed by calculating the degree of correspondence between key-frames of an unknown motion with the models in the database. The originality of this modelling is that the posture of a person is represented by a weighting silhouette representing the pose; weights which materialize variations of postures (As movement is executed in a different way from a person to

I. Bloch and R.M. Cesar, Jr. (Eds.): CIARP 2010, LNCS 6419, pp. 202–211, 2010.

another). The advantage of this approach is its simplicity and ease of processing and calculations. The selection of key positions is done automatically which is not always the case in other work [13].

2 System Overview

Human action is, in general, characterized by a sequence of specific body postures. So, the problem that we proposed to solve is to determine models representing these poses, to take them as references in order to recognize human action of everyday life with a fixed camera. The system implementation consists of three parts shown in figure 1. After the background subtraction, we have a human silhouette, then we centre this silhouette in a frame with predefined size, and as the last phase of preprocessing we select a set of frames (key-frames) that will represent the movement achieved in a video sequence. This last step, allows us to process just a fewer number of frames (the key-frames) instead of the entire input sequence frames. In modelling phase, we first calculate skeleton of silhouette in the key-frames. Then we use this skeleton as a starting point for building the model of the posture associated (weighting module). The out put of the modelling phase is a database of models representing movements (run, walk…). The last part of the implementation is the activity recognition module. This module use as input, a database cited above, and an unknown video sequence.

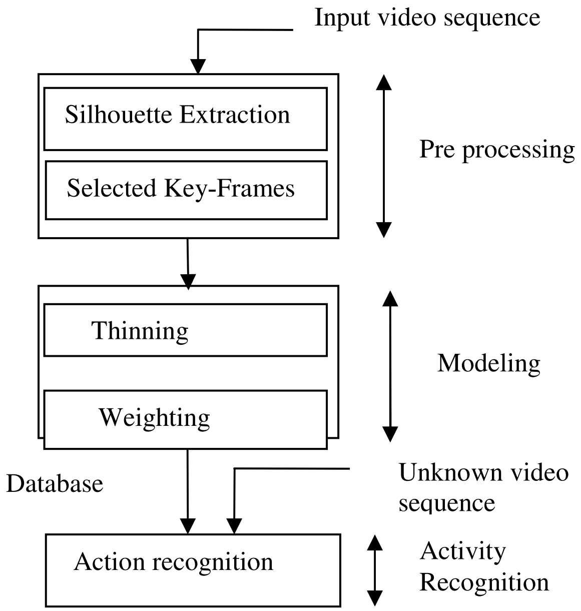


Fig. 1. Illustration of the processing stage of the system

3 Selected Key-Frames

An action is often described as a sequence of discrete postures. For determining which postures are the ones which can represent the movement, we treat the frames (given after the background subtraction) in pairs. This step is dependent on the accuracy of the tracking step (tracking process is not performed in this work) and is very important for the next process. So, we calculate the percentage of pixels (***perc***) that is different between two successive frames. Then we compare this value to a predefined threshold. If ***perc*** is upper than the threshold then the frame is selected to be a key-frame; otherwise it is not selected and we process the same treatment between the next frame and the last key-frame selected (See fig.2). The percentage is calculated as follows:

$$\text{Perc} = \text{Diff} / \text{Add} \tag{1}$$

Where **Diff** is the number of common pixels of two consecutives frames (given by the XOR operation between two frames).

Add is the number of all pixels in the two consecutives frames (given by the OR operation between two frames).

Perc is the percentage of pixels which is different between the two frames. This process allows us to quantify the difference in pixels using percentages to avoid relying on the number of image pixels (which change from one image to another).

This step allows us to define the keys postures (keys-frames) of a given movement. All these postures will therefore be selected to represent the movement. So, instead of processing the entire input sequence frame to recognize an unknown movement, we have to process just a fewer number of frames (the key-frames) (See figure 2.b). Note that this step can give us any number of key-frames in accordance with the velocity of the movement and the length sequence video.

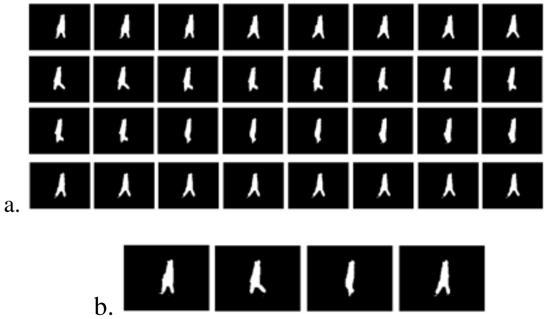


Fig. 2. Sequence of frames before and after selected key-frames. a. Input Sequence video, b. Selected key-frames.

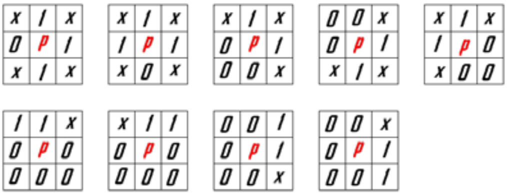
4 Modelling

The idea of creating weighted models of postures comes from the fact that a movement is executed by several people in a similar manner. Indeed, the different postures representing a movement for a given person are almost the same for any other person with slight deformations. These deformations are represented in models through the weights assigned to their pixels. The weight distribution in the frame model will be such that the skeleton pixels will receive the highest weight and distance from this position the more we will assign lower weights.

4.1 Building Models

The key-frames given by the postures selection step, on the input sequence video, are processed for building the models. The building models occur on two steps: thinning the silhouette then weighting the obtained skeleton.

Thinning. A skeleton is a geometric representation of an object in a dimension less than the input object. It can describe a compact way the properties of an object, especially its shape [19][20]. The algorithm we use is based on the topological thinning. The image analysis is to find simple points of the object of interest. To enjoy the benefits of parallel methods of thinning and conservation topological skeleton of sequential methods, we implement a hybrid algorithm. This algorithm is thinning the silhouette of two sides, north-east and south-west alternating direction at each iteration so as to obtain a skeleton centred in the image (See figure 3.b). We can divide the 16 cases of simple points in two groups: On one side the points which represent the north and/or east of the subject of interest. They are found in the following cases:



x	1	x
0	p	1
x	1	x

x	1	x
1	p	1
x	0	x

x	1	x
0	p	1
0	0	x

0	0	x
0	p	1
x	1	x

x	1	x
1	p	0
x	0	0

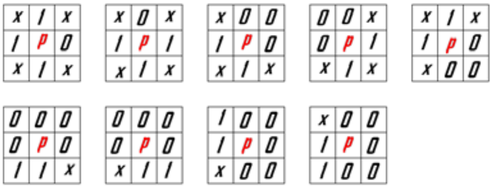
1	1	x
0	p	0
0	0	0

x	1	1
0	p	0
0	0	0

0	0	1
0	p	1
0	0	x

0	0	x
0	p	1
0	0	1

On the other side the points representing the south and / or west of the object of interest:



x	1	x
1	p	0
x	1	x

x	0	x
1	p	1
x	1	x

x	0	0
1	p	0
x	1	x

0	0	x
0	p	1
x	1	x

x	1	x
1	p	0
x	0	0

0	0	0
0	p	0
1	1	x

0	0	0
0	p	0
x	1	1

1	0	0
1	p	0
x	0	0

x	0	0
1	p	0
1	0	0

Some pixels may belong to both cases; this does not affect the course of the algorithm. The pixels removed are those located in opposite of the scanning direction of the image. This operation is repeated until no more simple point is detected.

The skeleton obtained is sometimes beyond the skeletal branches called "barbules". We call a branch; any set of pixels forming an eight-connected path whose elements

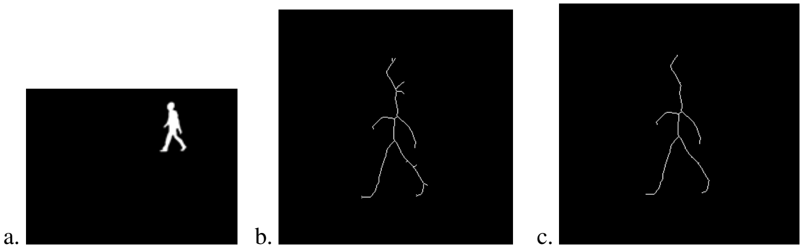


Fig. 3. Thinning a- Input silhouette, b- Skeleton, c-Skeleton after removing barbules

have strictly two neighbours (except for the two end pixels). Several criteria exist to remove the barbules (branches). The most used and easiest to implement is the size criterion. All arcs of the skeleton whose length is less than a given threshold are considered noise (barbules) and are removed. Several iterations are sometimes necessary (see figure 3).

Weighting. The weighting is a process of assigning weights, represented by symbols (Z, Y… in figure 4), to pixels in an image. These weights are used to specify the relative importance of each pixel compared to others. Weighting is used in the classification of postures, to calculate the degree of similarity (or correlation) between an unknown form and a model in the database.

For building models from the skeleton we process the following steps:

Step1: Distribution of maximum values of weighting on the skeleton.

Let Z be the maximum weight assigned to all pixels of the skeleton (see Figure 4.b).

Step 2: Second layer .

Each pixel 8-neighbor related to Z is associated with weighted value Y. The weight of Y is smaller than that of Z. Y values represent a layer covering the skeleton (see Figure 4.c).

Step3: Third layer (and more).

We repeat the "Step 2" with lower weight values (X, W,…) to the previous layer until we reach a thickness desired for shape (this processing is, always, done on the last layer obtained by the previous step).

The difference in values between each layer remains constant. A direct relationship between weight and number of layer is represented by:

$$NbL = Val\_max - Val\_min / Step \tag{2}$$

Where **NbL** is the number of layers, **Val_max** is maximum weight, **Val_min** is minimum weight, **Step:** difference between each weight of layer (step = Z – Y = Y – X…). These parameters are determined experimentally.

This treatment gives us almost the same gait (look) as that of the input posture. But the pixels of the image model are weighted (see figure 5).

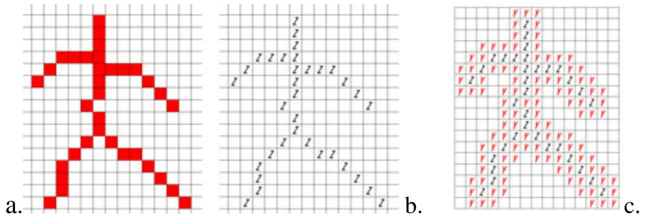


Fig. 4. Distribution of weights on the model a- Skeleton, b- Weights of the skeleton, c- Layer covering the skeleton

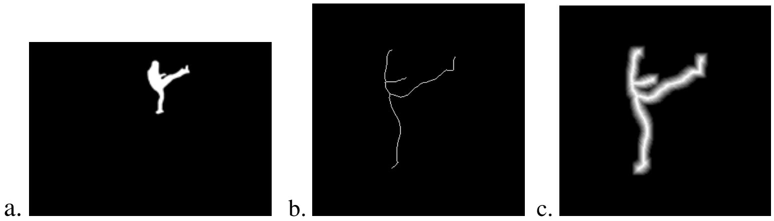


Fig. 5. A key-frame with his corresponding model, a-Input key-frame, b- Skeleton Key-frame, c- Corresponding model

4.2 Construction of Database of Models

The models are obtained after several treatments on selected images, namely: the normalization of size, thinning and weighting. For overlay models and skeleton, the frames must have the same size; a scaling is necessary. In this first work we are limited to process images that have relatively the same size. So we did not standardization of dimensions (scaling) on the images, but just add margins for the silhouette. We determine the endpoint of the object of interest for the four sides of the image. Then add columns and rows, on both sides of the object for getting a silhouette centred in an image of fixed dimensions.

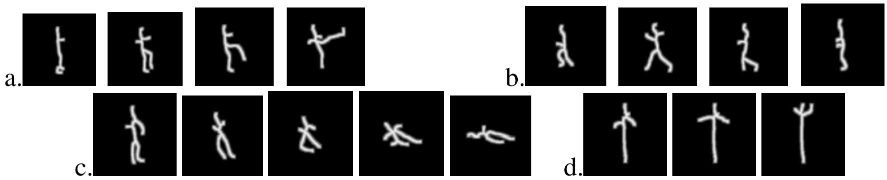


Fig. 6. Some models of the database, a. Models of ‘kick’, b. Models of ‘run’, c. Models of ‘Collapse Right’, d. Models of ‘hand wave’

From a video sequence of motion data, we select the key-frames representing the movement. Then, each of these key-frames undergoes treatment for scaling, thinning

and finally weighted. Our database has been constructed from seven input video representing different movements such as walk, run, punch, give a kick, collapse right, standup right and hand waving. Each movement is represented in the database by a set of image models of selected key-frames (see figure 6).
This first phase of the chain of recognition of human motion is, by analogy with other methods, the learning module.

5 Shape Matching

To validate the models we built, we propose to calculate a degree of correspondence between key-frames of unknown movement and the models in the database.

Input video sequences for an unknown movement is processed as be done for the sequence video which be used to build the database (select key-frames, scaling and thinning). We call degree of correspondence (***Deg_cor***) the sum of the weights of pixels in the model that overlap with the pixels of the skeleton of the unknown posture (figure 7.d.e). We calculate the number of pixels of the skeleton, in the unknown posture, and we multiply it by the maximum weight (Z), we obtain a value ***Val_max.*** We calculate then the Rate correspondence (***Rat_cor***) by:

$$\text{Rat\_cor} = (\text{Deg\_cor} \; / \; \text{Val\_max}) * 100 \tag{3}$$

The above calculation allows for comparison between two images: a model with an input image. But to make the recognition of a movement, we need to match a sequence of an unknown motion picture with a sequence of models (representing a movement) in the database.

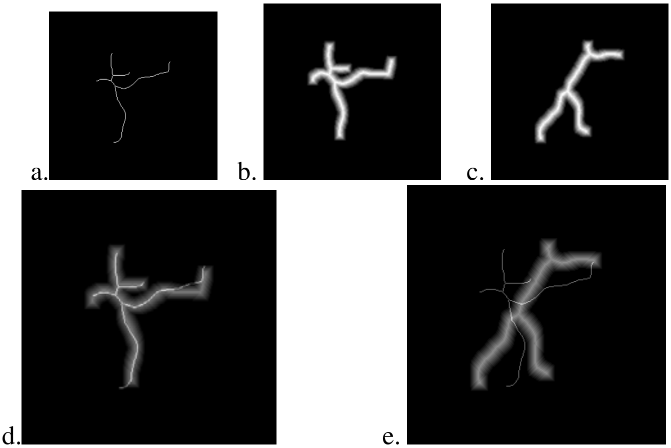
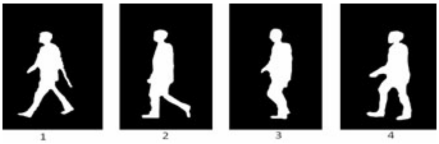


Fig. 7. Superimpose of skeleton on models. a- Skeleton of a posture, b- Corresponding Model, c- Non Corresponding model, d- superimpose of a skeleton on the corresponding model, e-superimpose of a skeleton on non corresponding model

6 Tests and Results

We have performed experiments on different video sequence actions. The system was trained using only one person for constructing the database. For the time being, the total number of activities in the database is seven (07). We give here an example of result obtained with an unknown input sequence.

We present (see figure 8.b) the results of the correlation calculated on an example of an unknown movement sequence "walking" (Four key postures: 1, 2, 3, 4; see figure 8.a) with the models of movements of the database "walk", "run" and "Kick". The first line of the matrix (respectively second, third and fourth) represents the correlations between the key position 1 (respectively 2, 3 and 4) movement to recognize with the different postures of the models.



a- Movement to recognize (with four key postures)

b-

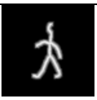
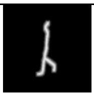
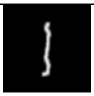
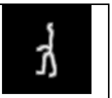
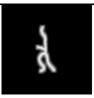
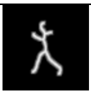
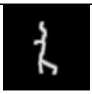
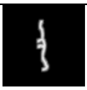
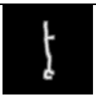
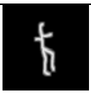
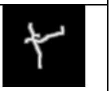
Walk				
1	**58.4746 %**	44.1702 %	54.5556 %	41.5484 %
2	22.7684 %	**66.8085 %**	30.5556 %	24.9677 %
3	40.8475 %	28.7660 %	**75 %**	38.7097 %
4	45.4237 %	42.9574 %	23.8889 %	**98.8390 %**
Run				
1	29.2181 %	28.0615 %	18.6517 %	34.4545 %
2	**53.4979 %**	**28.5164 %**	**55.8052 %**	42.6364 %
3	21.0700 %	15.1385 %	10.6367 %	**45.1818 %**
4	30.4527 %	19.0154 %	23.2210 %	21.4545 %
Kick				
1	27.9661 %	20.8833 %	17.5084 %	**24.7879 %**
2	35.7628 %	26.1199 %	**18.9226 %**	19.7576 %
3	**37.8814 %**	20.7571 %	10.7071 %	14 %
4	17.6271 %	**33.8170 %**	17.3064 %	16.9091 %

Fig. 8. The rats' correlation between an unknown sequence key frames with the models (walk, run and kick) in the database

From the results obtained, we can see that the degree of correspondence, between an unknown movement (figure 8.a) and his corresponding model, (figure 8.b, values in bold) give us higher values than those given for others models.

We used the video database given in [17] and our own sequence video. These first results encourage us to develop an approach for human motion recognition, which take, as a basis of knowledge, the models we built.

7 Conclusion

A novel modelling of human activities postures has been presented. The experiment, based on a simple compute of degree of correspondence shows encouraging results. For the future work, we envisage developing a recognition approach. Currently the implementation has some restrictions. The viewing direction is somewhat fixed and the background is assumed to be uniform making the segmentation of the silhouette easy. In addition, we assume that there is only one person in the field of view and that there is no occlusion. We plane to conduct more extensive tests to establish the limitation of our system.

References

1. Moeslund, T.B., Granum, E.: A Survey of computer vision-based human motion capture. Computer Vision and Image Understanding 81, 231–268 (2001)
2. Gavrila, D.M.: The visual analysis of human movement: a survey. Computer Vision and Image Understanding 73(1), 82–98 (1999), `http://www.idealibrary.com`
3. Aggarwal, J.K., Cai, O.: Human motion analysis: a review. Computer Vision and Image Understanding 73(3) (1999)
4. Bermermann, H.J.: Cybernetic functional and fuzzy sets. In: IEEE Systems, Man and Cybernetics Group Annual Symposium, pp. 248–254 (1971)
5. Sclaroff, S.: Deformable prototypes for encoding shape categories in image databases. Pattern Recognition 30(4), 627–640 (1996)
6. Belongie, S., Malik, J.: Matching with shape contexts. In: IEEE Workshop on Content-based Access of Image and Video Libraries (June 2000)
7. Carlson, S.: Order structure, correspondence and shape based categories. In: Forsyth, D., Mundy, J.L., Di Gesú, V., Cipolla, R. (eds.) Shape, Contour, and Grouping in Computer Vision. LNCS, vol. 1681, pp. 58–71. Springer, Heidelberg (1999)
8. Liu., T.L., Geiger, D.: Approximate tree matching and shape similarity. In: Proceeding of the IEEE International Conference on Computer Vision, Corfu, Greece, pp. 456–462 (1999)
9. Sharvit, D., Chan, J., Tek, H., Kimia, B.B.: Symmetry-based indexing of image database. J. Visual Commun. Image Representation 9(4), 366–380 (1998)
10. Siddiqi, K., Bouix, S., Tannenbaum, A., Zuker, S.W.: The Hamilton-Jacobi Skeleton. In: Proceeding of the IEEE International Conference on Computer Vision, Corfu, Greece, pp. 828–834 (1999)
11. Xie, J., Heng, P.A., Shaha, M.: Shape matching and modelling using skeletal context. Pattern Recognition 41, 1756–1767 (2008)
12. Goh, W.-B.: Strategies for shape matching using skeletons. Computer Vision and Image Understanding 110, 326–345 (2008)

13. Carlsonn, S., Sullivan, S.: Action Recognition by shape matching to Key Frame. In: Workshop on Models versus Exemplars in Computer Vision at CVPR (2001)
14. Kellokumpu, V., Pietikanen, M., Heikkila, J.: Human Activity Recognition Using Sequences of Postures. In: IAPR Conference on Machine Vision Applications (MVA 2005), Tsukuba Science City, Japan, pp. 570–573 (2005)
15. Huang, L., Wan, G., Liu, C.: An improved parallel thinning algorithm. In: ICDAR, vol. 02, p. 780 (2003)
16. Jang, B.K., Chin, R.T.: One-pass parallel thinning: analysis, properties, and quantitative evaluation. IEEE Transactions Pattern Analysis and Machine Intelligence 14(11), 1129–1140 (1992)
17. Schuldt, Laptev, Caputo: Proc. ICPR 2004, Cambridge, UK (2004)
18. Achard, C., Qu, X., Mokhber, A., Milgram, M.: A novel approach for recognition of human actions with semi-global features. Machine Vision and Applications 19, 27–34 (2008)
19. Ronse, C.: Pixel simple et nombres de Yokoi, LSIIT URM 7005 CNRS-ULP, Département d'informatique (2007),
    http://arthure.u-strasbg.fr/~ronse/TIDOC/index.html
20. Djahromi, A.K.: Binary Image Processing, Department of Electrical Engineering University of Texas at Arlington,
    http://www-ee.uta.edu/Online/Devarajan/ee6358/BinaryImage.pdf