Ghost: A Human Body Part Labeling System Using Silhouettes

Ismail Haritaoglu, David Harwood and Larry S. Davis
Computer Vision Laboratory, University of Maryland, USA
hismail,harwood,lsd@umiacs.umd.edu

Abstract

Ghost is a real time system for estimating human body posture and detecting body parts in monochromatic imagery. It constructs a silhouette based body model to determine the location of the body parts while people are in generic postures. It combines hierarchical body pose estimation, a convex hull analysis of the silhouette, and a partial mapping from the body parts to the silhouette segments using a distance transform method that does not violate the topology of the human body. Experimental results demonstrate robustness and real-time performance of the proposed algorithm.

1 Introduction

Visual interpretation of people and their movements is an important issue in many applications, such as surveillance systems, and virtual reality interfaces. The ability to find and track people's body and body parts is therefore an important visual problem. In this work, we describe a monocular system, *Ghost*, to find and label people's body parts from images in real time.

Ghost is a real time system for detecting human body parts in monochromatic imagery. It constructs a silhouette-based body model to determine the location of the body parts. In this paper, we describe the computational models employed by *Ghost* to predict the location of the six main body parts (e.g, the head, hands(2), feet(2), and the torso) while a person is in any of a number of postures.

There has been recent interest on detecting , tracking and understanding human actions [13, 11, 7, 5, 3, 2]. With increasing processor power, more attention has been given to intelligent real time surveillance systems which detect people [13, 7, 5, 6, 4] and recognize their activities [2, 11]. Detecting and tracking human body parts (the head, hands, feet) is important in understanding human activities. When the system detects a person in the scene, a geometric body part analysis is applied to the foreground regions where the person is detected. W^4 [5, 6] and other previous systems [8, 7, 13, 9, 1] make the assumption that people perform actions while they are in an upright-standing posture, and perform their body part analysis based on that assumption. However, a surveillance systems should operate and track body parts while the person is in other generic postures (e.g., sitting, crawling). We propose an algorithm that works not only in the upright-standing posture but also in other generic postures. A hierarchical posture representation is proposed in *Ghost*. Any body posture is classified a *main posture* (in this paper we consider four main postures: standing, sitting, crawling-bending, and laying down) and then each main posture is sub-classified into one of three *view-based appearances* (front-view, left-side, and right-side).

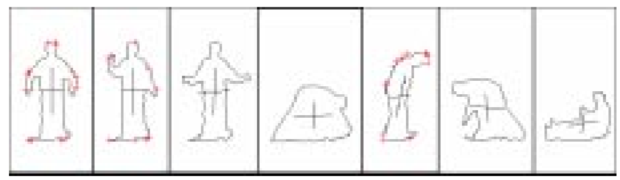

Figure 1: Some silhouette boundaries of the same person in different posture

Ghost works under the control of W^4. W^4 invokes *Ghost* when it needs to locate body parts (when W^4 detects a new person or when it loses a body part which was being tracked). *Ghost* locates the body parts and passes their estimated locations to W^4. W^4 then tracks the individual body parts using its correlation method.

Our system is motivated by two basic observations on the relative location of body parts while people are in action.

- It is very likely that the head, hands, elbows, feet, and knees lie on the silhouette boundary.
- The human body in any given posture has a topological structure which constrains the relative locations of body parts. The order of body parts along the silhouette boundary does not typically change when people perform ax action while maintaining a generic posture (walking); however, the order does change when they change their generic posture (walking to sitting).

Ghost uses a silhouette-based body model which consists of 6 primary body parts (head, hands(2), feet(2), and torso), which we want to locate, and 10 secondary parts(elbows(2), knees(2), shoulders(2), armpits(2), hip, and upper back) which could be on the silhouette boundary and can help to locate the primary parts using the topology of the human body. The outline of the algorithm used in *Ghost* is as follows:

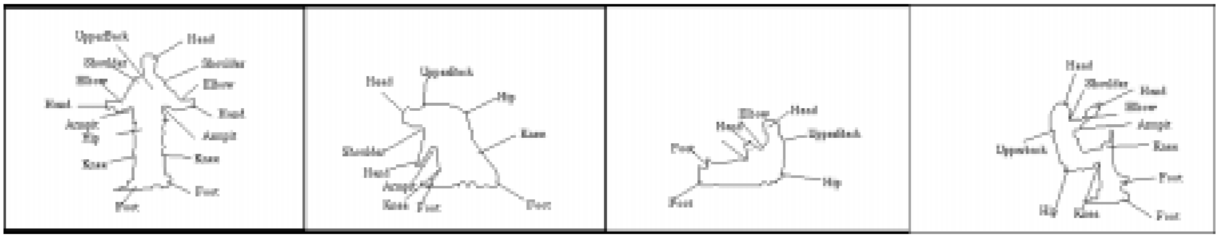

Figure 2: Examples of the order of the body parts on the silhouette boundary

1. A hierarchical body posture analysis is applied to the silhouette to compute the similarities of horizontal and vertical projection histograms of the detected silhouette and the main postures. The body posture which yields the highest similarity measure is taken as the estimated posture.
2. A recursive convex-hull algorithm is applied to find possible body part locations on the silhouette boundary.
3. The location of the head is predicted using the major axis of the silhouette, the hull vertices, and the topology of the estimated body posture.
4. When the head location is determined, a topological analysis is applied to eliminate the hull vertices which won't be labeled as body parts, and to map the remaining hull vertices to the body parts using a topological-order preserved distance transform calculation.

2 Related Work

Pfinder [13] is a real-time system for tracking a person which uses a multi-class statistical model of color and shape to segment the person from a background scene. It locates and tracks a person's head and hands under a wide range of viewing condition. Iwasawa[7] introduced a real-time method to estimate the posture of a human from thermal images acquired by an infrared camera regardless of the background and lighting conditions. Lueugh and Yang [9] tried to solve the body labeling problem in a sequence of images. They describe a segmentation method, using difference images and past history. Akita [1] used a key frame sequence of stick figures to approximately predict the locations of body parts in intermediate frames.

3 Detection of Human Body Parts

Ghost takes silhouettes of people as input to locate body parts. Silhouettes are segmented from the background by a four stage process: thresholding, noise cleaning, morphological filtering and object detection. A detailed description of our background scene modeling and foreground region detection can be found in [5, 6].

3.1 2D body modeling using silhouettes

Ghost uses a silhouette-based body model which consists of 6 primary body parts (head, hands, feet, and torso), which we want to locate, and 10 secondary parts(elbows, knees, shoulders, armpits, hip, and upper back) which could be on the silhouette boundary and can help in locating the primary parts using the topology of the human body (Figure 2). It is not necessary to find all body parts in any single frame; some of them might not be visible from a given point of view.

The primary and secondary body parts should be consistent with the order of the respective main posture (with small variations). These orders are preserved as long as the body stays in the same main posture. For example, if we start from the head (Figure 2) in clockwise order, the main order for the upright-standing pose is head-shoulder-elbow-hand-armpit-knee-foot-foot-knee-armpit-hand-elbow-shoulder-head. This order could vary for different view points. Some parts could be missing on the silhouette boundary or some parts are switched in the order locally (elbow-hand or hand-elbow) because of relative motion of the parts or local occlusions. However, the relative location of some parts (head, feet) should be preserved. For example, head-elbow-shoulder or hand-feet-knee are unacceptable partial orders for the standing posture. Any order of the body parts in the given silhouette should be generated from the order of the main posture by deleting the missing parts or switching the location of some neighbor parts (elbow-hand). Therefore, if we know the posture of the given silhouette and the location of at least one body part, the labeling problem becomes one of mapping the set of body parts to the set of the silhouette segments without violating the expected order of the respective posture.

3.2 Estimation of the human body posture

People can be in many different postures while they are performing actions. Each posture has different appearances (views) varying with the point of view. Our system makes the assumption that the angle between the view direction and the ground plane is within 0^o to $+60^o$. We collected examples of people over a wide range of views, and extracted their silhouettes to discover the order of body parts on the silhouette for different postures. We observed that four different main postures (standing, sitting, crawling-bending and laying down) have large differences in the order of

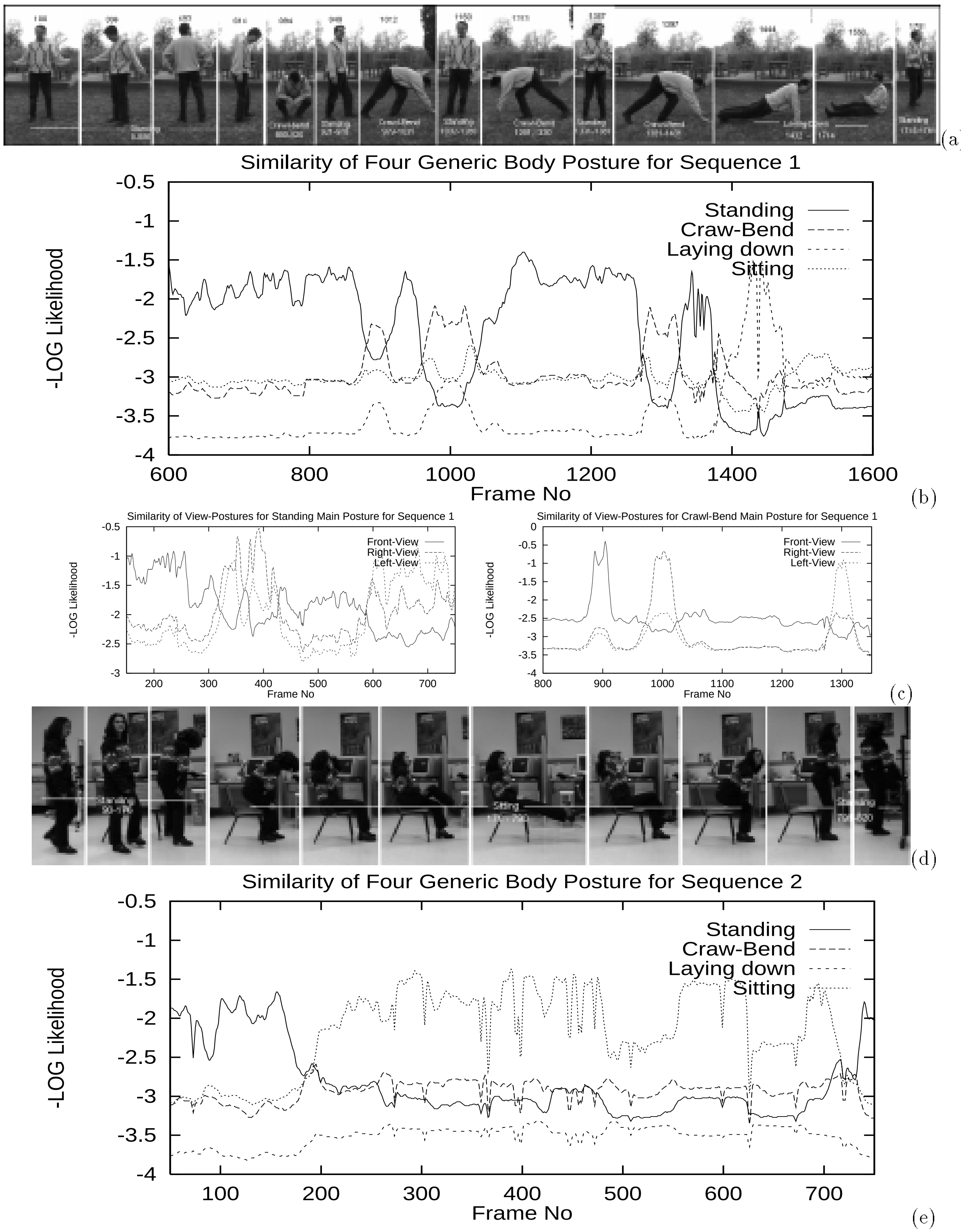

Figure 3: The similarity of four main postures for two different sequences

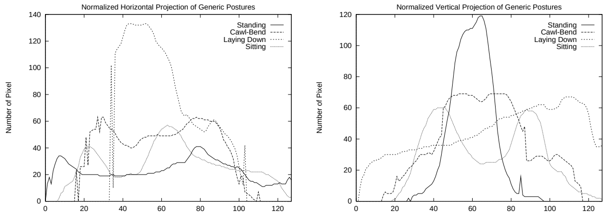

Figure 4: The vertical and horizontal normalized projections of standing, crawling-bending and laying down postures used in the body posture estimation

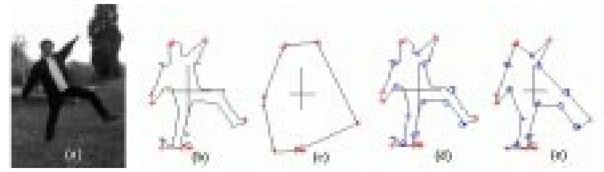

Figure 5: The body in the standing pose (a), the convex-hull vertices (b), the shape approximation by the convex-hull vertices (c), the convex and concave hull vertices (d), and the shape approximation by convex and concave hull vertices (e).

body parts. The order in other postures is typically a variation of one of the main postures. *Ghost* classifies the observed human body posture in a *hierarchical* manner. Any body posture is classified into one of the four *main postures* (standing, sitting, crawling-bending, laying) and then each main posture is classified into one of three *view-based appearances* (front-view, left-side view and right-side view).

A body posture is represented by the normalized horizontal and vertical projection histograms (projection), the median coordinate, and the major axis of its silhouette. Average normalized horizontal and vertical projection templates for each main posture (and for each view-based appearance of each main posture) were computed experimentally using 4500 silhouettes of 7 different people in 3 different views. These features are used to determine the similarity of the given posture to one of the four main postures. In Figure 4, the normalized vertical and horizontal projection templates of the standing, crawling-bending, laying down, and sitting postures used in the body posture estimation in *Ghost* are shown.

Ghost determines and normalize the vertical and horizontal projections of the silhouette. Normalization is done by rescaling the silhouette into a fixed vertical length while keeping the original aspect ratio. *Ghost* then compares the observed silhouette with the projection templates of the four main postures using the sum of absolute difference method to estimate the most similar main posture. Let S^i be the similarity between the detected silhouette and ith main posture, H^i and V^i the horizontal and vertical projections of ith main posture,respectively, and P and R the horizontal and vertical projections of the detected silhouette. S_i is calculated as:

$$S_i = -log(\sum_{h}^{128}\sum_{v}^{128} |(H_h^i - P_h)| + |(V_v^i - R_v)| \quad (1)$$

Ghost determines the most similar main posture by using the highest score; it then applies the same method to determine the most similar view-based appearance for the estimated main posture. In Figure 3, the result of main posture and the view-based appearance estimation are shown for two sequences. In sequence 1 (1750 frames), the person performs some simple work-out actions. He was in the following postures (with frame numbers): standing (0-850), crawling-bending (850-910), standing (920-970), crawling-bending(left view) (970-1080), standing (1080-1270), crawling-bending (right view) (1270-1320), standing (1320-1380) and laying down (1400-1470). The graph in Figure 3(b) shows how the classification method is able to select the correct posture over time. 95% of the postures were successfully classified in that sequence. Figure 3(c) shows the view-based appearance estimation for the standing (left) and crawling-bending (right) main postures for sequence 1. Note that when the body is in the crawling-bending posture (the peaks in the figure 3(c)-right), the view-based appearance was successfully classified. In sequence 2 (750 frame), the person performs a "walk-sit-walk" action. She was in the following postures: standing (0-180), sitting (200-700) and standing (710-750). The graph in Figure 3(e) shows classification results for sequence 2. 98% of the postures are correctly classified in that sequence.

3.3 Detection of the convex and concave hulls on silhouettes

Because of the topology of humans, it is likely that some body parts always appear on the extreme points or curvature maxima of the silhouette boundary. We need to find those points on the silhouette in order to find the set of locations which will be labeled as body parts. We implemented a recursive convex hull algorithm (Graham scan) to find these vertices (hulls) which are candidates to belong to one of the primary body parts. We modified the Graham scan convex-hull algorithm to be able to exploit the silhouette property and speed up the calculations. Convex hull vertices alone are not enough for shape representation of the silhouette. We also need also some concave hull vertices to obtain a better representation of the silhouette. Therefore, we implemented our hull analysis in recursive: we first find the convex hull vertices of the silhouette in the first iteration, and then we applied the same convex hull algorithm to each of the silhouette segments between two convex hull vertices detected in first iteration to find the concave hull vertices. Figure 5 shows one of the standing poses (a), the convex hull vertices on its silhouette(b), the shape approximation by the convex hull vertices alone (c), the convex and concave hull vertices on its silhouette(d), and the shape approximation by convex and concave hull vertices(e).

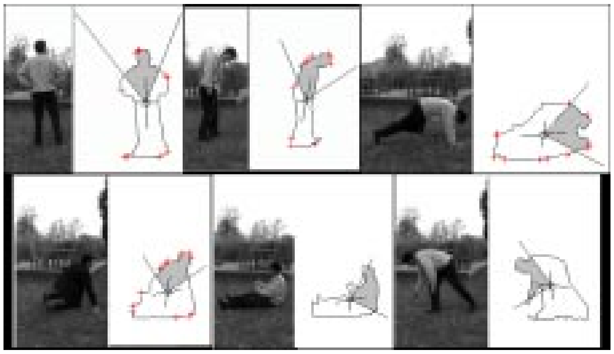

Figure 6: the silhouette segments for the head location predicted by *Ghost*

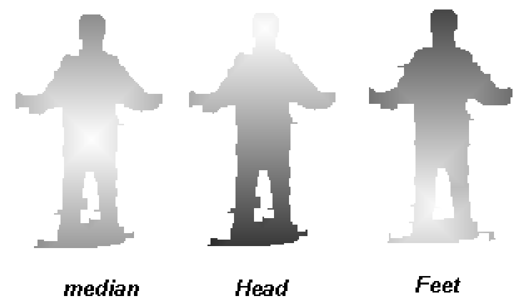

Figure 7: Distance transform from the head, feet, and median

3.4 Prediction of body part locations

3.4.1 Prediction of the head location

We need to select one body part as a starting point to be able use the order constraints and relative distances of body parts to find the other. We take the head as a reference point to locate the other parts. The head is a stable body part compared to the others, and its location can be easily predicted. *Ghost* tries to find a silhouette segment which includes the head by combining the constraints on the order of body parts on the silhouette for the expected main posture, the principal axis of the silhouette and the ground plane information. Let p be major axis of the silhouette, and let l_1 and l_2 be the two lines which intersect p at the median coordinate of the silhouette. The angles between l_1 and p is α^o and the angle between l_2 and p is $-\alpha^o$. *Ghost* determines the silhouette segments whose starting and end points intersect points between the silhouette and lines l_1 and l_2, respectively. α varies between 22^o to 45^o according to the estimated posture and the aspect ratio of the silhouette. The ground plane is used to eliminate the silhouette segment which is on the opposite side of the head with respect to the median coordinate. In Figure 6, the silhouette segments for the predicted head locations are shown.

3.4.2 Prediction of the feet and hands and torso locations

After *Ghost* determines the head location, it tries to find the other primary body parts in the order of the feet, the hands, and the torso using prior-knowledge about the topology of the estimated body posture. Let M^i be a subset of the primary and secondary body parts which are predicted to be visible given the hypnotized main posture. Let C^t be set of the convex and concave hull vertices detected on the silhouette boundary. We need to find a fast and simple partial mapping from M^i to C^t to solve the labeling problem; the labeling should be consistent with the order of body parts for the estimated posture. Relative path distances are used to check if the mapping is consistent with the order. The relative distances and the order-constraints for each main posture and view-based appearance of the main posture are calculated experimentally. Initially, the path distances from the head and median to hull vertices are computed using a distance transform method. After the initial distance calculation is done, each hull vertex is assigned to one or more body parts according to their relative distances as long as they satisfy the order-constraints of the estimated posture. The hull vertices which are assigned to the feet are labeled first, then the hull vertices for shoulders, upper backs, hands, elbows, armpits and knees are labeled. If there is more than one vertex assigned to the same body part, another search is applied by narrowing the order constraints for those vertices. While labeling a vertex, the distances to the previously labeled vertices should also be consistent. The torso is located between the median coordinate and the head along the major axis of the silhouette. Examples of body part labeling results of *Ghost* are shown in Figure 8

4 Conclusion and Discussion

We have described an approach to detect body parts of people using silhouettes that is implemented in a a system called *Ghost* which is working under the control of the human detection and tracking system W^4 [5]. *Ghost* has been implemented in C++ and runs under the Windows NT operating system. Currently, for 320x240 resolution gray scale images, *Ghost* runs at 10-20 Hz on a PC which has dual 200 Mhz pentium processor. It has the capability to detect the body parts (the head, hands, feet, and torso) of multiple people in complex backgrounds. Figure 8 illustrates the body part labeling results for different postures.

There are several directions that we are pursuing to improve the performance of *Ghost* and to extend its capabilities. We are studying the use of a statistical model to formulate the partial mapping from the hulls on the boundary to the body part. The relative motion of the body parts could be employed to get more robust and correct mappings. In current implementation, shadows create problems in locating body parts which are too close to the ground. Better re-

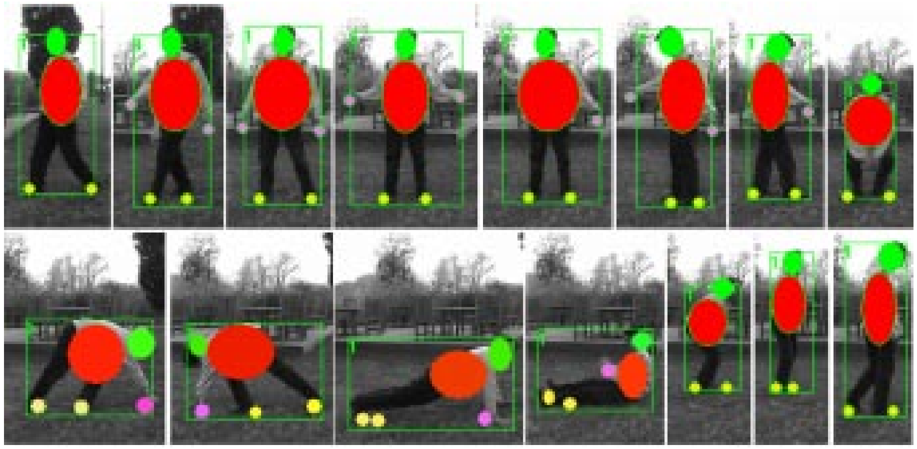

Figure 8: Examples of using silhouette model to locate the body parts in different actions

sults could be obtained if shadows are removed from the silhouette, perhaps based on stereo or motion.

5 Acknowledgments

This work was supported by DARPA contract VSAM DAAL0197K0102

References

[1] K. Akita. image sequence analysis of real world human motion. *Pattern Recognition*, 17(4):73–83, 1984.

[2] A. Bobick and J. Davis. Real-time recognition of activity using temporal templates. In *IEEE Workshop on Application of Computer Vision*, pages 1233–1251, 1996.

[3] A. Bobick, J. D. S., Intille, F. Baird, L. Cambell, Y. Irinov, C. Pinhanez, and A. Wilson. Kidsroom: Action recognition in an interactive story environment. Technical Report 398, M.I.T Perceptual Computing, 1996.

[4] Q. Chai and J. K. Aggarwal. Tracking human motion using multiple cameras. In *International Conference on Patern Recognition*, pages 68–72, 1996.

[5] I. Haritaoglu, D. Harwood, and L. Davis. W4: Who, when, where, what: A real time system for detecting and tracking people. In *In Proc. of Third Face and Gesture Recognition Conf.*, pp:222-227 April 1998

[6] I. Haritaoglu, D. Harwood, and L. Davis. W4S: A real time system for detecting and tracking people in 2.5 D In *In Proc. of ECCV*, June 1998

[7] S. Iwasawa, K. Ebihara, J. Ohya, and S. Morishima. Real-time estimation of human body posture from monocular thermal images. In *Proceedings of the Computer Vision and Pattern Recognition*, 1997.

[8] S. X Ju, M. J. Black, Y. Yacoob. Cardboard people: A parameterized model of articulated image motion. In *Second International Conference on Automatic Face and Gesture Recognition*, Killington, Vermont, 1996.

[9] M. Leung and Y. H. Yang. A region based approach for human body motion analysis. *Pattern Recognition*, 20(3):321–339, 1987.

[10] T. Olson and F. Brill. Moving object detection and event recognition algorithms for smart cameras. *Proc. DARPA Image Understanding Workshop*, pages 159–175, 1997.

[11] R. Polana and R. Nelson. Recognizing activities. In *International Conference on Patern Recognition*, 1994.

[12] R. Polona and R. Nelson. Low level recognition of human motion. In *Non Rigid Motion Workshop*, pages 21–29, 1994.

[13] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *In Proc. of the SPIE Conference on Integration Issues in Large Commercial Media Delivery Systems*, October 1995.

[14] C. Wren and A. Pentland. Dynamic modeling of human motion. *In Proc. of Third Face and Gesture Recognition Conf.*, pp:22-27 April 1998