Human Activity Analysis Based on a Torso-less Representation

Jen-Hui Chuang, Chun-Wei Lee, Kuo-Hua Lo
Dept. of Computer Science, National Chiao Tung University, Hsinchu 300, Taiwan
monoz.cs95g@nctu.edu.tw

Abstract

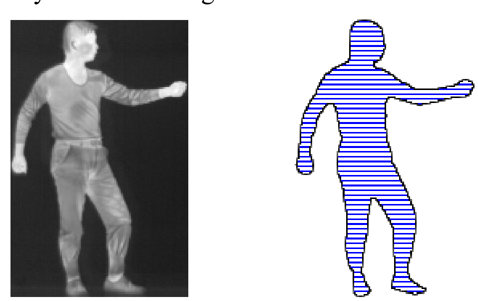
Human activity recognition is a popular topic in the field of computer vision. While most analysis algorithms take into consideration of the whole human body, the movements of merely the head and limbs are often informative enough in many practical applications. In this paper, a novel approach is proposed to track these extruding parts of a human body in consecutive images. Accordingly, a simplified torso-less pattern of gesture is proposed to represent human activities, and with its effectiveness verified subjectively by some experimental results. Such a representation can not only ensure the privacy of the person being watched, but is also suitable for real-time surveillance based on bandwidth-limited communication since only a very small amount of data are used compared to conventional approaches.

1. Introduction

Human activity analysis has received increasing attention from researchers in the fields of image processing and computer vision during the past few years. Related research topics include human detection, motion tracking, posture recognition, and behavior analysis from image sequences. One common approach is to build a model of human body so that the associated activity analysis is carried out with respect to an underlying model [1-5]. However, intensive computations are usually required for a complex model. Researchers adopt different approaches to avoid resorting to such models. In [6] and [7], the human gait is recognized by analyzing silhouette of human images. In [8], the skeleton structure of human region is extracted before human behaviors are recognized by an efficient eigenspace-based method. In [9], the projection histogram of binary image is used as one of the features to discriminate different postures.

While different types and amounts of image data are used, with or without a model of human body, the torso of human body often plays an essential role in the above approaches to human activity analysis. Nevertheless, it is observed that human activities can often be understood roughly from the movements of the head and limbs. Therefore, identification of these extruding parts of a human body, which is the main purpose of this paper, may already provide sufficient information in many applications, e.g., video surveillance. Skeletal representations, e.g., the central axis presented in [1], are often adopted to describe their shapes since human head and limbs are roughly elongated in shape.

In this paper, the above shapes are identified by generating their approximate medial axes (MA). The approach is an extension of that proposed in [10] where similar shape description is derived to facilitate subsequent classification of a human chromosome. Unlike [1], the approach will generate the axes without assuming that an initial axis point is given. In this paper, infrared (IR) images obtained from a night vision system, like the one shown in Fig. 1 (a), are considered for simplicity. The approach will work equally well for other image sources if the human body can be separated reasonably from the background.



(a) (b)

Figure 1. (a) An IR image. (b) Cross-sections obtained from horizontal scan-lines.

978-1-4244-2175-6/08/$25.00

While automatic activity analysis using the above data of human head and limbs alone is possible and is under investigation, a simple torso-less representation of human, which can be examined by a person, is proposed. Such a representation could not only ensure the privacy of the person being watched, but is also suitable for real-time surveillance based on bandwidth-limited communication since only a very small amount of data, e.g., the ends of the above axes, are used.

2. Identification of Head and Limbs

In [10], approximate medial axis of a chromosome is obtained by connecting midpoints of *selected* cross-sections obtained from four equally spaced angular orientations. Since the chromosome has elongated shape with nearly uniform width, the majority of these cross-sections will have lengths close to its width and can be selected easily using a length histogram. In this section, the above approach of shape recognition is generalized, with necessary modifications, to take into the account the following properties of a human body: (i) The body has branches, (ii) Each of its extruding parts, e.g., limbs and head, has a different elongated shape with graduated changing widths.

2.1 Generation of Cross-sections

For the identification of head/limbs regions of a human body, we first generate its cross-sections along scan-lines in four angular orientations: 0°, 45°, 90°, and 135°, as in [10]. The idea is to identify appropriate cross-sections for each of these regions, whose shape and orientation are initially unknown, so that subsequent processes can connect midpoints of these cross-sections to form an approximate MA of that region. Fig. 1(b) shows cross-sections thus obtained for Fig. 1(a) with horizontal scan-lines. One can see that these cross-sections represent the orientation and the shape of the head and one foot very well.

To save computation time, instead of scanning all image pixels in an image, an equally spaced subset of scan lines can be used by considering only one in every δ scan lines, shown in Fig. 2(a). However, if δ is too big, the approximate MA obtained by subsequent procedures may not represent a curved shape properly. For our experiments with 320×240 IR images, $\delta = 3$ is chosen as a trade-off between computing speed and appropriate shape representation. (δ can actually be determined automatically with respect to the average dimension of a human image.) Finally, cross-sections with length less than 5 pixels are assumed to be noise and are discarded.

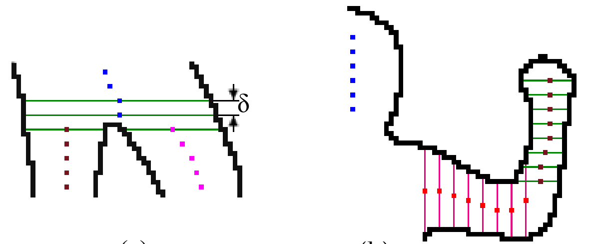


Figure 2. (a) A close-up view of Fig. 1(a). (b) Two MA segments to be connected.

2.2 Derivation of Candidate of MA Segments

Without using the length histogram of the cross-sections obtained in the previous subsection, as in [10], candidates of MA segments are derived in a slightly different fashion according to the aforementioned properties (i) and (ii). Here, midpoints of *all* cross-sections obtained in the previous subsection are considered as possible samples along MA segments. For each of the four scan-line directions, two midpoints of consecutive cross-sections will be connected into a candidate segment of MA if (A) their distance is smaller than a threshold T_A, and (B) the length difference of the two cross-sections is less than T_B. Fig. 2(a) shows a branch point of human body where the midpoints are not connected so that a total of three MA segments will be derived.

In the rest of the paper, the analysis results are based on $T_A = 2\delta$ for (A) and $T_B = 5$ pixels for (B). While the latter can be regarded as a tolerance of noise as that considered in the previous subsection, for two consecutive cross-sections of the same length, the former corresponds to an intersection angle of 153.4° (compared to a perfect 90°) between the MA segment and the cross-sections.

2.3 Screening for Appropriate MA Segments

After the process mentioned above, four sets of the MA segments are obtained from four scan-line directions and more than one MA segments may be generated for each part of the human body. Since very short MA segments will not represent an elongated object well, those with less than three midpoints will be removed. On the other hand, the longest MA segment can best represent an elongated object; therefore, if two or more MA segments intersect, the longest MA segment is chosen while all the shorter ones are removed. Moreover, short MA segment with wide cross-sections will also be removed, since they will not be parts of an elongated object.

2.4 Forming of approximate MAs

The MA segments derived above may need to be linked into an approximate MA, since multiple segments may be derived for each of the limbs. Fig. 2(b) shows the generated MA segments of an arm and a gripping fist, obtained from two scan-line directions, after the above screening. Apparently, they should be linked to represent the arm's shape. Let d denote the distance between the closest end points of two MA segments, MA_i and MA_j, and L_{av} denote the average length of the cross-sections of an MA segment. In our approach, if $d < \delta$, or $\delta < d_{ij} < 2\delta$ and $| L_{av}^{i} - L_{av}^{j} | < 2\delta$, the two MA segments will be linked into one. Fig. 3(a) shows some approximate MAs thus obtained.

3. A torso-less representation of human activity

For a complete skeletal representation of the extruding parts of the human body, the two ends of each approximate MA, the proximal end and the distal one, will need to be discriminated. While the location of the former is supposed to be near to the connecting point to the torso and may vary significantly in different image frames, the location of the latter can often be found with little deviation. Thus, the identification and tracking of the above distal ends, which will then give the extruding ends of the head and limbs, will be discussed next. A simplified torso-less pattern using mainly such ends is suggested in the next section to demonstrate some experimental results.

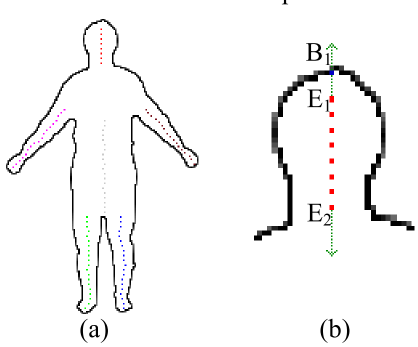


Figure 3. (a) Some approximate MAs. (b) Finding the extruding end of the head.

3.1 Identification of extruding ends of a person

Consider the approximate MA of the head shown in Fig. 3(b) with E_1 and E_2 being its two ends. Let E_1E_2 intersect the boundary of human body at B_1 and B_2, respectively. If $E_1B_1 < E_2B_2$, then E_1 can be identified easily as the distal end; otherwise E_2 is the distal end. Subsequently, simple rules regarding the relative length of an MA and the associated L_{av}, which are omitted for brevity, can be applied for the discrimination among head, hands and feet for the distal ends identified above. Besides, further simplifications of the process identifying these extruding ends are possible if additional information about their relative locations available, e.g., through a gesture known a priori. Similarly, for a sequence of images, two consecutive images may share their identification results, as discussed next as a tracking problem.

3.2 Tracking the movements of limbs and head

To analyze human activity involving continuous changes in gesture, movements in different parts of the human body may need to be identified. Nonetheless, tracking the movements of only limbs and head usually suffice the need. On the other hand, it is easy to see that the tracking will also simplify the identification of these body parts. In our approach, a simple bounding box, which is constructed from the MA's two ends and L_{av}, is used to roughly represent the head and each of the limbs in an image. For a sufficiently high processing speed, movements of each body part between two consecutive images will be limited. Therefore, if B_{t-1} and B_t are obtained from image frames t-1 and t, respectively, and

$$\frac{B_{t-1} \cap B_t}{min(B_{t-1}, B_t)} > TH\_S \qquad (1)$$

for some threshold TH_S, the two bounding boxes are assumed to be representing the same body part. The minimum value is used in (1) to allow significant size and shape change of a tracked region, as confirmed in the following experimental results.

4. Experimental Results

In order to demonstrate the effectiveness of the skeletal representation derived above, a simplified torso-less pattern is proposed here. A virtual torso having a length three times of the head MA and a width twice of the head L_{av} is first established. The extruding ends of human body found in Sec. 3 are then connected to fixed locations of this torso. While the feet are connected using parallelograms, ellipses are used for head/hands. All these simple shapes have widths proportional to corresponding L_{av}'s. For each occluded part, a default shape pattern is depicted.

Figs. 4 and 5 show two sequences of the above patterns obtained for walking and sitting, respectively.

Simulation is performed under Windows XP, 1 GB RAM, 1.5G Intel Centrino CPU, and with processing speed over 20 frames per second. In order to assess the usefulness of pattern sequences thus generated, six sequences of simple human activities, including walking, sitting, raising hands, and lifting up a leg, are presented to thirty users for gesture recognition. An average recognition rate of more than 90% suggests that these torso-less patterns can indeed provided sufficient information of human activity in many occasions.

5. Summary

In this paper, we proposed an effective method of analyzing human activity based on tracking the movements of only the head and limbs. A simplified torso-less pattern of gesture is developed accordingly as well. Experimental results show that some human activities can indeed be comprehended quite accurately by the proposed approach. Since using such a pattern will resolve the privacy issue, the efficiency of the approach makes it very suitable for various real-time surveillance applications. Extensions of the proposed approach to automatic analysis of human activity are also under investigation.

Acknowledgements

This research is supported by the National Science Council under grant NSC93-2218-E-009-030.

References

[1] B. Fan and Z.-F. Wang, "Pose Estimation of Human Body Based on Silhouette Images," *IEEE Int. Conf. on Information Acquisition*, pp. 296-300, 2004.

[2] M. K. H. Leung and Y.-H. Yang, "First Sight: A Human Body Outline Labeling System," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 17, No.4, pp. 359–377, 1995.

[3] C. Hu, Q. Yu, Y. Li, and S. Ma, "Extraction of Parametric Human Model for Posture Recognition Using Genetic Algorithm," *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 518 – 523, 2000.

[4] C.-C. Chen, J.-W. Hsieh, and Y.-Y. Huang, "Segmentation of Human Body Parts Using Deformable Triangulation," *IEEE 18th Int. Conf. on Pattern Recognition*, Vol. 1, pp. 355-358, 2006.

[5] S. Kim, C.-B. Park, and S.-W. Lee, "Tracking 3D Human Body using Particle Filter in Moving Monocular Camera," *IEEE 18th Int. Conf. on Pattern Recognition*, Vol. 4, pp. 805-808, 2006.

[6] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette Analysis-based Gait Recognition for Human Identification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 25, No. 12, pp. 1505-1518, 2003.

[7] H. Fujiyoshi and A. J. Lipton. "Real-Time Human Motion Analysis by Image Skeletonization,"*4th IEEE Workshop on Applications of Computer Vision*, pp. 15-21, 1998.

[8] Y. Yuan, Z. Miao, and S. Hu, "Real-Time Human Behavior Recognition in Intelligent Environment," *IEEE 8th Int. Conf. on Signal Processing*, Vol. 3, pp. 16-20, 2006.

[9] P. Guo and Z. Miao, "Projection Histogram Based Human Posture Recognition," *IEEE the 8th Int. Conf. on Signal Processing*, Vol2, pp. 16-20, 2006.

[10] J.-H. Kao, J.-H. Chuang, and T. Wang, "Chromosome Classification Based on The Band Profile Similarity along Approximate Medial Axis," *Pattern Recognition*, Vol. 41, No. 1, pp. 77-89, 2008.

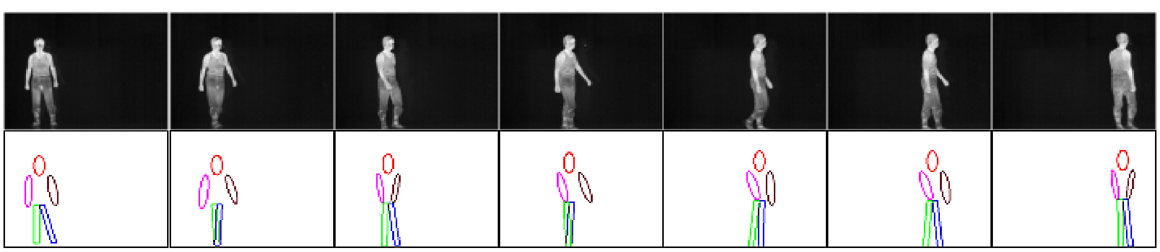


Figure 4. Torso-less patterns obtained for *walking*.

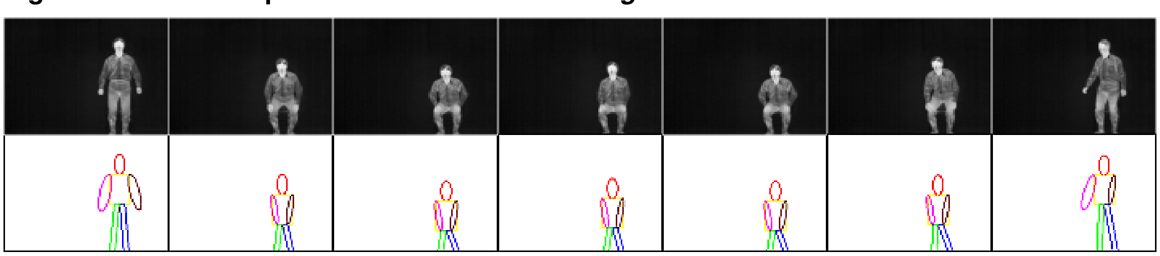


Figure 5. Torso-less patterns obtained for *sitting*.