

LABELING OF HUMAN BODY PARTS ¹

Shyamsundar Rajaram, Jezekiel Ben-Arie and Dayan Manohar Sivalingam

ECE Department M/C 154, University Of Illinois at Chicago,
851 S.Morgan St., Chicago, IL 60607, USA
phone/fax: 001-312-996-2648 e-mail: benarie@ece.uic.edu

Abstract: *We develop a novel method for human body parts labeling, given a set of body parts that are represented by bars in the image. First, we assume that in addition to the head and torso, the human body has 8 major body parts namely, the upper and lower arms (left and right) and the upper and lower legs (left and right) and all these parts can be approximated by cylinders that appear as rectangular bars in the image. One of the original contributions of this paper is the novel use of the spatial distributions of the location and orientation of the bars employed to find a correct role assignment for each bar by maximizing a joint probability function. Thus, we formulate the problem of labeling human body parts as an optimal hypothesis selection problem where each hypothesis corresponds to a particular role assignment. Next, we develop an algorithm which identifies the optimal hypothesis. We test our method with 80 images that depict human poses and obtain correct labeling in 98.7% of the trials. We also test our method in cases of bar detection errors, occlusion and noisy images and obtain quite good results.*

Key words: *Human Body Part Labeling, Optimal Hypothesis Estimation, Spatial Distributions of Torso Center*

1. INTRODUCTION

In this research, we develop a novel method for human body part labeling using their spatial locations and orientations. The problem of human body part labeling is a very important step in recognizing human activities from video. The labeling is necessary as a preliminary step in tracking body parts for activity recognition [1]. This problem has applications in fields such as human-machine interaction, ergonomics, kinesiology, sports, security surveillance and in virtual environments. The main challenge in this problem is that humans can take a multitude of poses ranging from a simple walking pose to complicated poses like that of baseball pitching. In each pose, one has to detect, locate and identify all the major body parts for tracking or activity recognition.

In this work, we limit the labeling to human poses captured only in a range of views that are approximately frontal. The problem is inherently a very difficult problem because there are 8 different major body parts and there are hardly any appearance based cues that help to label them as particular body parts. In this work, we make use of the spatial orientation of the body parts as a main cue to identify the best possible labeling. Thus, the problem of labeling is formulated as a hypothesis probability estimation algorithm and we develop a theoretical framework in section 2 that serves as the basis for our labeling approach.

The basic premise of our work is the modeling of human body as a collection of parts arranged in a deformable configuration. This premise has been used in several previous works [7] [3] [4] [5][8]. However, our approach differs considerably from these works in the method we use to find the labeling. In [7], Haritaoglu et al. use the silhouette of a human performing the activity and capture the convex and concave points on the boundary of

¹This work was supported by the National Science Foundation (NSF) Grants No. IIS-0208300, IIS-9876904 and IIS-9979774.

the silhouette. Then, they perform labeling of the parts starting from the head which is assumed as the reference location by using path constraints. In [3], Hogg tracks a walking person by comparing edge projections of a model with the edges in an image. The best fit defines the configuration of a human. Kakadiaris et al. [6] develop a novel technique for human body part identification by treating the human as a single object to start with and then split the object into new objects as the human body deforms by moving. The end result is that each limb is represented by an object. Felzenszwalb et al.[8] develop a statistical framework for object recognition using pictorial structures. Using the statistical technique, they train models to represent human bodies and use these models to locate the different body parts in images. We find that our method is entirely different from all these approaches.

We use a synthetic human body model during the learning phase of our method, to estimate the relative location of the center of the torso with respect to each of the body parts. Each body part in the synthetic human model is maneuvered to assume all possible orientations. The relative locations of the torso with respect to the center of that body part are stored in one single image, where the center of the image corresponds to the center of the body part. The spatial distributions of the center of the torso are assumed to be uniform distributions. This assumption, that is based on the premise that each orientation of the body part is equally likely, simplifies the overall computation. The next step in our work involves assignment of roles to the detected bars (that represent body parts) in a given image. A novel method is used in identifying the optimal role assignment of the bars out of all the possible role assignments. For each combination of role assignments, we accumulate the spatial distributions of the center of the torso relative to the center of each of the bars constituting the combination. One of the novel contributions of our method is in identifying the optimal assignment as the assignment which has the **maximum overlap** of these spatial distributions. The development of the spatial distributions and the labeling algorithm will be further elaborated in the section 3. In section 2, we describe the theory behind our approach. Section 4 discusses the application of our approach and demonstrates the method experimentally.

2. THEORETICAL FOUNDATIONS OF OUR APPROACH

In this section, we describe the theoretical foundation to our approach in identifying human body parts from a set of detected bars. In this work, we approximate the image of human body by 8 bars which represent the 2 upper arms, the 2 lower arms, the 2 upper legs and the 2 lower legs. Let the bars that are detected in the image by the bar detection routine ² be represented by $B = \{b_1, b_2 \dots b_n\}$ where n is the number of bars detected by the routine and we denote the total number of possible roles of each bar by m and the roles themselves by $R = \{r_0, r_1 \dots r_m\}$. Since, we have eight bar type body parts $m = 8$. In addition, we define an empty role r_0 that does not indicate any of the 8 possible roles and is used to indicate if a bar is to be denoted as an out-lier.

The problem here is to assign a set of roles to a set of detected bars. The problem of body part detection can be solved by formulating it as an optimal hypothesis selection problem. A hypothesis is represented as a n -tuple $H = \{h_1, h_2 \dots h_n\}$, where if $h_i = j$, it means that the bar b_i is assigned the role r_j where j can take values in the range $[0 \dots m]$. We assume that the bars are independent of each other and that it is equally likely for any bar to assume any role. Another assumption is that all the pixels in the image L_1, L_2, \dots, L_c (c is the number of pixels in the image) have the same a priori probability of being identified as a central location of the torso. As we shall elaborate later in section 3, the location of the

²We developed a bar detection routine that is based on signatures of Gabor filters[2]. However, one may develop other approaches for bar detection as well.

torso is estimated when a bar assumes a specified role. We estimate the best assignment by choosing the hypothesis H , which has the maximum probability $P(H)$. $P(H)$ can be written as

$$P(H) = P(\{h_1, h_2, \dots, h_n\}) \quad (1)$$

and it can be further reduced by using law of total probability as,

$$\begin{aligned} P(H) &= P(\{h_1, h_2, \dots, h_n\}) \\ &= P(\{h_1, h_2, \dots, h_n\}|L_1)P(L_1) + P(\{h_1, h_2, \dots, h_n\}|L_2)P(L_2) + \dots + \\ &\quad P(\{h_1, h_2, \dots, h_n\}|L_c)P(L_c) \end{aligned} \quad (2)$$

Since, we assume that the roles assigned to bars are all mutually statistically independent of each other,

$$P(\{h_1, h_2, \dots, h_n\}|L_1) = P(B_1^{h_1}|L_1)P(B_2^{h_2}|L_1) \dots P(B_n^{h_n}|L_1) \quad (3)$$

where $B_i^{h_j}$ denotes the assignment of role h_j to the bar B_i for all i, j . Further, $P(L_i)$ for all $i = 1 \dots c$ is some fixed constant based on our assumption that all pixels are equally probable of being the center of the torso. Hence, the more the locations that support the hypothesis H , the larger is $P(H)$. We assume that the correct assignment is composed from body parts that fit better and therefore it has larger overlap and thus higher probability. Hence, we find the optimal hypothesis H_o by maximizing the function given by Eq. (2) and Eq. (3).

$$H_o = \underset{H}{\operatorname{arg\,max}} P(H) \quad (4)$$

We further elaborate the process of finding the optimal hypothesis in section 3.

3. BODY PART LABELING AND DISTRIBUTIONS

In this work, we are interested in the frontal view of upright human poses. We use a synthetic human body model M for developing these spatial distributions. For each of the different upper body part roles say the left upper arm, we develop spatial distributions of possible locations of the center of the torso by maneuvering the synthetic human body model M such that the left upper arm assumes all feasible orientations and the location of the center of the torso relative to the left upper arm is noted for each such orientation. A result of such a step is as shown in Fig. 1a which shows the possible locations of the center of the torso for different orientations of the left upper arm. Similarly, the spatial distribution is developed for all the upper body parts and the distributions are as shown in Fig. 1 To simplify the computations, we assume that all of the orientations are equally likely and hence all the spatial distributions are uniform. For the lower body parts denoted by X , we note that location of X is dependent on its orientation and also on the orientation of the upper body part denoted by Y which is directly connected to it. Hence, while we find the spatial distributions for X , we find the relative location of the center of the torso with respect to the center of X , by maneuvering M in such a way that all the possible locations of X are considered by rotating the upper body part Y and further rotate the lower body part X with respect to Y . In such a way, we find the spatial distribution for all the lower body parts.

When a human performs an activity in front of a camera, if the pose of the body part is not parallel to the camera plane, then the part is foreshortened in its projection onto

the 2D image. To overcome the problem of foreshortening and errors in the bar detection stage, while estimating the spatial distributions, we consider various body part lengths in the range of $l/2$ to $2 * l$ where l is the standard length of the body part. Further, to allow for measurement noise the location of the center of the torso is also considered as a rectangular region instead of just a point in the image.

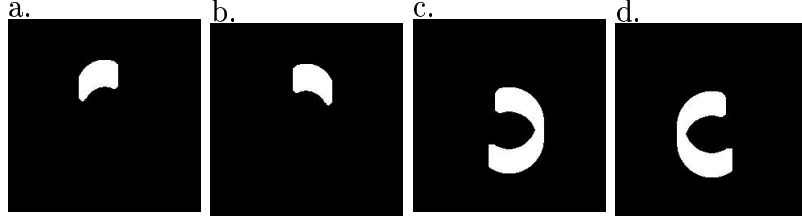


Fig. 1. Spatial distribution of the center of torso for (a)- various orientations of the right upper leg where the center of the image represents the center of the right upper leg, (b)- various orientations of the left upper leg where the center of the image represents the center of the left upper leg, (c)- various orientations of the right upper arm where the center of the image represents the center of the right upper arm, (d)- various orientations of the left upper arm where the center of the image represents the center of the left upper arm.

Apart from the relative location of the center of the torso, we find the orientation of the torso with respect to the upper body parts while computing the spatial distribution. We keep track of this information to eliminate parts of the spatial distribution since the distribution includes relative location of the center of the torso for all possible orientations of the upper body part and as we know the orientation of the upper body part denoted by θ , we consider only parts of the spatial distribution which were formed with the orientation θ .

The objective of this work is to identify the correct assignment of roles to major body parts that are represented in the image by bars. As mentioned in the introduction, this is quite a difficult problem since the detected bars do not lend cues which can be positively used to identify them as either arms or legs. So, we use the spatial location and orientation of the bars to identify their roles. If there are b detected bars and r roles then the total number of assignments possible is $b!/(b-r)!$ if $b > r$ or $r!/(r-b)!$ if $b < r$. We have to find the assignment which is the correct assignment out of all these assignments by maximizing the probability function given by Eq. (2) and Eq. (3).

The spatial distributions denote the probability distributions of location of the center of the torso relative to a given location of an upper body part in the spatial sense. Since, it is assumed that all orientations of any body part are equally likely, the distributions are uniform. The first stage of the labeling process is to accumulate votes for all the $b * r$ possible role assignments and keeping track of the bar and the role that resulted in a vote. So, we have a four dimensional array where the first two dimensions represent the dimensionality of the image, the third dimension identifies the bar and the fourth dimension refers to the role of the bar. For any pixel location say L_i , the vote corresponding to bar j when it is assigned role k is equivalent to $P(B_j^k | L_i)$ which is one of the constituents of the right hand side of Eq. (3).

In the second stage, we consider each possible labeling combination H and estimate the probability $P(H)$ given by Eq. (2) and Eq. (3). Each combination constitutes a hypothesis and the hypothesis having the highest function value is chosen to be the optimal hypothesis. This is the hypothesis that is used for labeling the different bars with appropriate labels. Intuitively, the above stage can be explained as identifying the assignment

which produces the maximum overlap of the spatial distributions which are accumulated as votes in the first stage.

In order to decrease the number of computations involved we do an initial screening of the pixels. We calculate the sum of all the votes received by every pixel in the image. Let the maximum value obtained over all the pixels be denoted by MAX . During the labeling process, only those pixels are considered whose sum of votes is greater than $\alpha * MAX$ where α is some threshold in the range $[0 \dots 1]$. This step significantly decreases the computations of our labeling algorithm. We observe here that the region of pixels resulting from the initial screening are all concentrated around the center of the torso and this region includes all the pixels that belong to the optimal assignment combination.

4. EXPERIMENTAL RESULTS

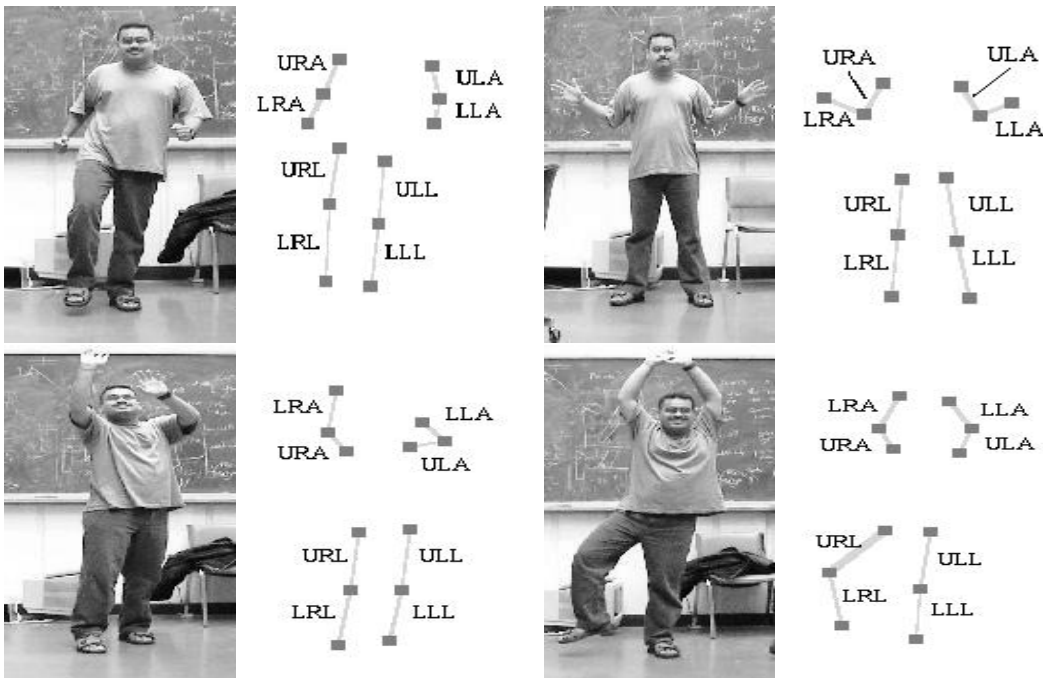


Fig. 2. Human poses and labeling performed by our system. All the labelings are correct.

In this work, we use a novel bar detection scheme developed as part of the thesis work done under Prof. Jezekiel Ben-Arie by Manoj [2] in our lab. The bar detection algorithm is based on extracting Gabor signatures of different kinds of bars and to search for similar signatures in the image where humans are present. More information about the bar detection algorithm can be found in [2].

Results of body part labeling are shown in Fig. 2. The method gives 98.7% correct recognition for 80 different human poses. As seen from the results, our algorithm works even in the case of strong foreshortening of body parts. Our algorithm's inherent robustness to foreshortening is because we allow variation of the body part lengths when we estimate the spatial distributions. This also helps in overcoming sizable errors in the bar detection algorithm.

Further, we tested our method in cases where some bars were missed due to occlusion or due to noise in the image. Our system achieves 98.7% correct labeling when no body

parts are missing, 97% correct labeling when 1 body part is missing, 94% correct labeling when 2 body parts are missing, 89% when 3 body parts are missing and 84% when 4 body parts are missing. Because of the larger spatial distribution of the lower body parts when compared to that of the upper body parts, the method makes the occasional error of labeling a bar which is actually the upper body part as a lower body part when the actual bar corresponding to the lower body part is missing.

5. CONCLUSION

In this work, we present a novel approach for human body part labeling by a part based model for representing humans. We randomly assign roles to different bars in the image. The hypothesis which has the highest probability of being the optimal assignment of roles is selected. The method gives us 98.7% correct labeling for 80 different human poses. In our approach, we represent the human body using a part based model in the approximate frontal view and therefore our method is restricted to a range of frontal views of humans (approximately ± 45 degrees in azimuth). However, this problem can be solved in general by constructing more elaborate spatial distributions which accommodate sideviews as well.

REFERENCES

- [1] Ben-Arie, J., Wang, Z., Pandit, P. and Rajaram, S., "Human Activity Recognition using Multidimensional Indexing," **IEEE Transactions in Pattern Analysis and Machine Intelligence**, Vol. 24, No. 8, August 2002 .
- [2] Seshadrinathan, M., "Face detection and Recognition," **Masters' thesis**, ECE Dept. University of Illinois at Chicago, Fall 2002.
- [3] D. Hogg, "Model-based Vision: a Program to See a Walking Person" *Image and Vision Computing*, Vol 1, No. 1, pp. 5-20, 1983.
- [4] A. Pentland, "Recognition by Parts" *Proc. Int'l Conf on Computer Vision*, pp. 612-620, 1987.
- [5] J. O'Rourke and N. I. Badler, "Model-Based Image Analysis of Human Motion using Constraint Propagation," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, Vol 2, No. 6, pp. 522-536, Nov. 1980.
- [6] I. A. Kakadiaris and D. Metaxas, "3D Human Body Model Acquisition from Multiple Views", *International Conference on Computer Vision*, pp. 618-623, Cambridge, MA, June 20-23, 1995.
- [7] Ismail Haritaoglu, David Harwood and Larry S. Davis, "W⁴: Real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, August, 2000, pp 809-830.
- [8] Pedro F. Felzenszwalb, "Object Recognition with Pictorial Structures," *Masters thesis at MIT*, May 2001.