# Recursive Context Reasoning for Human Detection and Parts Identification

Liang Zhao and Chuck Thorpe
The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
Email: LZhao@cs.cmu.edu


## Abstract

*Human detection and body parts identification are important and challenging problems in computer vision. High performance human detection depends on reliable contour extraction, but contour extraction is an under constrained problem without the knowledge about the objects to be detected. This paper proposes a recursive context reasoning (RCR) approach to solving the above dilemma. A TRS[1]-invariant probabilistic model is designed to encode the shapes of the body parts and the context information — the size and spatial relationships between body parts. A Bayesian framework is developed to perform human detection and part identification under partial occlusion. A contour reconstruction procedure is introduced to integrate the human model and the identified body parts to predict the shapes and locations of the parts missed by the contour detector; the refined contours are used to reevaluate the likelihood ratio. Therefore, contour extraction, part identification, and human detection are improved iteratively. The experimental results of the RCR approach to human detection and body parts identification in cluttered scenes are very encouraging.*


## 1 Introduction

Human detection and body parts identification are important and challenging problems in computer vision. They have a wealth of applications ranging from automated navigation [24] to human-computer interaction [23]. The main challenge facing a vision-based human detector is the high degree of variability with the human appearance due to articulated motion, partial occlusion, and inconsistent cloth texture. Contours (the silhouettes of objects) are the common features used to overcome inconsistent texture; parts-based approaches [1, 3, 4, 5, 6, 9] can handle occlusion and articulated motion effectively. Therefore, we employ the contours or the outlines of human body parts as features to detect people.

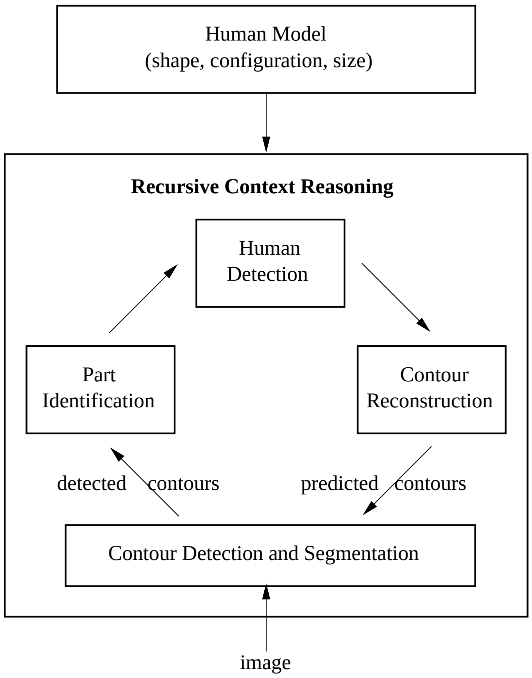


Figure 1: Flow chart of the RCR algorithm

Many part-based methods [1, 3, 4, 5, 6, 7, 18, 9] have been proposed but only with limited success. Some approaches [1, 2] assumed that the object parts are statistically independent in order to simplify the problem. This assumption works when the object parts have very salient shapes (such as face features). However, it loses the capability to check the consistency between the object parts by ignoring the spatial and size relationships between the parts. Another assumption made by the previous work [4, 6] is that the object contours have been correctly detected and par-

[1] TRS is the abbreviation of translation, rotation, and scaling

titioned before performing part identification and human detection. However, contour detection and partition are very unstable procedures and are unable to provide perfect and complete contours due to occlusion, the cluttered background, or the low contrast.

In this paper, we propose a recursive context reasoning approach (as illustrated in Fig. 1) to perform and refine contour extraction, human detection, and parts identification iteratively. A TRS[2]-invariant probabilistic model is designed to encode the shapes of the body parts and the context information — the size and spatial relationships between body parts. A Bayesian framework is developed to perform human detection and part identification under partial occlusion. A contour reconstruction procedure is introduced to integrate the human model and the identified body parts through a Kalman filter [11] to predict the shapes and locations of the parts missed by the contour detector; the refined contours are used to reevaluate the likelihood ratio. Therefore, contour extraction, part identification, and human detection are improved iteratively.

Burl *et al.* [8] have proposed a related method which combines the intensity pattern and the spatial relationship between the face features to detect faces from the cluttered environment. However, they do not use the spatial relationship to help detect face features, and no size relationship and recursive procedure is involved in face detection. The main reason is that face features have very distinctive patterns and can be detected reliably based on local features. In human detection, we rely heavily on the spatial and size relationships to identify body parts, because human features like arms and legs do not present very distinctive texture patterns or shape cues.

The remainder of this paper is organized as follows. Section 2 presents the problem formulation in the Bayesian framework. Section 3 gives the outline of the RCR algorithm. The TRS-invariant probabilistic human model is introduced in Section 4. The details and the experiments of part identification, human detection, and human contour reconstruction are given in Section 5 to 7, respectively. Section 8 discusses the contributions and the limitations of the proposed algorithm, and proposes the future work.

## 2 Problem Formulation

### 2.1 Human Detection

We employ the contours (the silhouettes) of objects) as the features for human detection and formulate the human detection and part identification problems in a Bayesian framework. The human detection problem is: given a detected contour $C$, determine whether this contour represents the silhouette of a person (the hypothesis $w$) or not (the hypothesis $\bar{w}$).

The optimal method to detect a person is based on the maximum *a posteriori* (MAP) rule: If $P(w|C) \geq P(\bar{w}|C)$, select $w$, i.e. a person is detected. Otherwise, select $\bar{w}$, i.e., no person is detected.

By applying Bayes' rule, we can rewrite the optimal decision rule as follows:

$$\frac{P(C|w)}{P(C|\bar{w})} \underset{\bar{w}}{\overset{w}{\gtrless}} \frac{P(\bar{w})}{P(w)} \qquad (1)$$

In Eq. (1), the left hand side is the likelihood ratio; the right hand side is the ratio of prior probabilities that can be considered as a threshold controlling the receiver operating characteristics (ROC) of the detector.

Because the detected contours are not always complete and perfect, there are situations when the information provided by the detected contours is not enough to make the final decision. Therefore, We set two thresholds to define an uncertainty region $[\lambda_1, \lambda_2]$. Then the decision rule becomes:

$$\begin{cases} \text{select } w, & \text{if } P(C|w)/P(C|\bar{w}) > \lambda_2 \\ \text{no decision}, & \text{if } P(C|w)/P(C|\bar{w}) \in [\lambda_1, \lambda_2] \\ \text{select } \bar{w}, & \text{if } P(C|w)/P(C|\bar{w}) < \lambda_1 \end{cases} \qquad (2)$$

### 2.2 Parts Identification

The parts identification problem is to match a set of partitioned contour segments to a set of human body parts. We formulate this problem as an optimal hypothesis selection problem. Assume that the contour $C$ is partitioned into $n$ segments $C = \{c_1, c_2, ..., c_n\}$, and that the human model consists of $m$ body parts or human features: $F = \{f_1, f_2, ..., f_m\}$. Let $H = (h_1, h_2, ..., h_m, view)$ represent a hypothesis about the matching between the contour segments in $C$ and the human features in $F$, and about the viewpoint from which the person is observed, where

$$h_i = \begin{cases} j, & \text{if } f_i \text{ is matched with } c_j \\ 0, & \text{if } f_i \text{ is occluded or not detected} \end{cases}$$

and $view \in \{front/back, side\}$. This is not a one to one mapping. It allows certain body parts being occluded and also allows certain contour segments being outliers (not labeled by any body parts). Unlike previous parts identification approaches that either assume no outliers [16] or developed two procedures [6, 5] to locate body parts and attached objects independently.

[2] TRS is the abbreviation of translation, rotation, and scaling

Our approach provides a single procedure to achieve both goals.

We select the MAP hypothesis $H^*$ from the hypothesis space $\mathcal{H}$, s.t.

$$H^* = arg \max_H P(H|C,w) \quad (3)$$
$$= arg \max_H P(C|H,w)P(H|w). \quad (4)$$

We define

$$G(H) \triangleq P(C|H,w)P(H|w) \quad (5)$$

as the *goodness function* that rates hypotheses.

## 3 Outline of the RCR Algorithm

The formal description of the RCR algorithm (as shown in Fig. 1) is as follows. Let $\hat{C}_t$ be a set of contour segments detected or reconstructed at step $t$.

**Step 1**: part identification: generating a hypothesis $H_t$ from $\hat{C}_t$, s.t. $H_t = arg \max_H G(H)$.

**Step 2**: human detection: calculating the likelihood ratio $\Lambda_t = P(\hat{c}_t|w)/P(\hat{c}_t|\bar{w})$ to determine if a person is present in the image based on the decision rule given in Eq. (2). If $\Lambda_t \in [\lambda_1, \lambda_2]$, then goto Step 3 to continue searching for a more complete and accurate set of contours. Otherwise, declaring that a person is detected if $\Lambda_t > \lambda_2$, or declaring that no person is detected if $\Lambda_t < \lambda_1$.

**Step 3**: Reconstructing the outlines of the body parts $\hat{C}^-_{t+1}$ from the outlines of the identified body parts $\hat{C}_t$ and the human model.

**Step 4**: Aligning $\hat{C}^-_{t+1}$ to the edge features in the image and generating the aligned contour set $\hat{C}^+_{t+1}$.

The stop conditions are (a) we can make a reliable decision on whether the image contains a person or not; or (b) no more human body parts can be identified from $\hat{C}_t$; or (c) there is no further improvement or changes to $\hat{C}_t$.

## 4 Human Body Model

The human body model plays an important role in the RCR algorithm. It provides the constraints for human detection and part identification, and for inferring the shapes and locations of the missed body parts from the detected body parts. For the purpose of human detection and model learning, we developed a TRS-invariant human body model. For the purpose of modeling the shape variations between different people, we encode the model parameters as joint probability distributions. We call the resulting model — *the TRS-invariant probabilistic model*. In the following subsections, we will describe the human body model in detail.

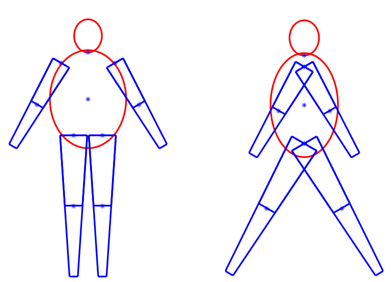


Figure 2: Human body model

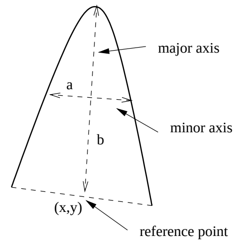


Figure 3: Body part model

### 4.1 TRS-Invariant Body Model

There are two 2D-human body models — the front-view model and the side-view model as shown in Fig. 2. Each human body model consists of six main body parts (the head, the torso, two arms and two legs) and eight subparts of the four limbs. The body parts are constrained to connect to each other at the joints. The body parts are modeled with ribbons [12, 13] as shown in Fig. 3. We define the axis or the spine of a ribbon as the major axis of a body part, the line segment that perpendicularly bisects the major axis as the minor axis. Then we define the ratio of the length of the major axis ($b$) to the length of the minor axis($a$) as the aspect ratio of a body part. The aspect ratio captures the global shape of a body part and is TRS-invariant. Each limb has a unique reference point which is located at the joint point (as shown with stars in Fig. 2). The reference point of the torso is located at its center. A body part model is parameterized with a vector $(s, b, x, y, \theta)$, where $s = a/b$ is the aspect ratio, $(x, y)$ are the coordinates of the body part's reference point, and $\theta$ is the orientation of the major axis. The body part model constrains the global shape of a body part while allowing local shape deformation.

Accordingly, the human body model is parameterized with four model matrixes: the shape vector $S = \{s_1, ..., s_m\}$, the size ratio matrix $R = \{r_{ij}\}, i, j = 1...m$, where $r_{ij} = b_i/b_j$, the spatial relationships or the configuration vector $X = (x_1, y_1, ..., x_m, y_m)$, and the orientation or the posture vector $\Theta = \{\theta_1, ..., \theta_m\}$. Obviously, $S$ and $R$ are TRS-invariant.

The position of a subpart $f_i$ is constrained by the length and the orientation of its main body part $f_j$:

$$(x_i, y_i) = (x_j, y_j) + b_j * (cos\theta_j, sin\theta_j). \quad (6)$$

Thus, we will just model the configuration of the six main body parts. Let the vector $X = (x_1, y_1, ..., x_6, y_6)$. To make this vector TRS-invariant, we map the coordinates of the joints to the normalized torso coordinate system with the length of the torso normalized to be 1. Then we obtain the TRS-invariant configuration model: $U = (0, 0, u_2, v_2, ..., u_6, v_6)$, where $(u_i, v_i) = ((x_i, y_i) - (x_1, y_1))/b_1$, $(x_1, y_1)$ are the coordinates of the torso's center, and $b_1$ is the length of the torso.

Currently, the human model is parameterized with three TRS-invariant matrixes: $S, R, U$, which constrain the shapes, the relative sizes and the positions of the body parts, respectively. In the future, we are going to learn the posture model in order to make stronger constraints on the appearance of a person.

### 4.2 TRS-Invariant Probabilistic Model

We make the human body model deformable by representing the model parameters with joint Gaussian distributions. The aspect ratio, the length, and the location of a body part are statistically independent of each other, thus, the three model matrixes $S, R, U$ are also statistically independent of each other. We model their probability distributions separately:

$$S \sim \mathcal{N}(\bar{S}, \Sigma_S) \quad (7)$$

$$R \sim \mathcal{N}(\bar{R}, \Sigma_R) \quad (8)$$

$$U \sim \mathcal{N}(\bar{U}, \Sigma_U). \quad (9)$$

The above probability distributions provide metrics to evaluate the shape, size relationship, and configuration similarities between the detected contour and the human body model. They are estimated from the measurements given by [15]. [15] provides both the body measurements of people at different ages and the clothing corrections.

## 5 Part Identification

In Section 2, we formulate the part identification problem as an optimal hypothesis selection problem. The selection of the best hypothesis is based on the goodness function $G(H)$ as defined in Eq.(5). In the following sections, we will describe the estimation of the goodness function, and the efficient hypothesis selection strategy in detail. The experimental results are given in Subsection 5.3.

### 5.1 Goodness Function Estimation

In order to calculate the goodness function $G(H)$, we need to estimate the likelihood $P(C|H, w)$ and the prior probability $P(H|w)$.

Let $\hat{S}, \hat{U}, \hat{R}$ be the model matrixes estimated from the identified body parts. Then we have $P(C|H, w) = P(\hat{S}, \hat{U}, \hat{R}|H, w)$. Because $S, U, R$ are mutually statistically independent, we can rewrite the likelihood $P(C|H, w)$ as

$$P(C|H, w) = P(\hat{S}|H, w)P(\hat{U}|H, w)P(\hat{R}|H, w) \quad (10)$$

The likelihoods $P(\hat{S}|H, w)$, $P(\hat{U}|H, w)$, and $P(\hat{R}|H, w)$ are estimated from Eq.(7-9), respectively. The likelihood $P(C|H, w)$ derived above constrains the best match between the detected contour segments and the human body parts to be of similar size relationships, and configurations as well as similar shapes.

The prior probability $P(H|w)$ is estimated based on the visibility and the detect-ability of human features. Assume that each candidate has the same probability to be the real human feature, then $P(H|w)$ can be written as

$$P(H|w) = \frac{P(d, view|w)}{\prod m_i} \quad (11)$$

$$= \frac{P(d|view, w)P(view|w)}{\prod m_i} \quad (12)$$

where $m_i$ is the number of candidates for a human feature $f_i$, $d = (d_1, d_2, ..., d_m)$ is a vector with $d_i = 0$, if $h_i = 0$, and $d_i = 1$, if $h_i \neq 0$. Both $P(d|view, w)$ and $P(view|w)$ can be estimated from the training data.

### 5.2 Optimal Hypothesis Selection

The exhaustive search of the best hypothesis $H^*$ is to check every possible hypothesis and find the best one based on the goodness function $G(H)$. However, the hypothesis space is generally very large due to combination explosion. In this paper, we developed a two-step search procedure to identify the body parts. First, we identify the main body parts based on the goodness function. Then, we derive the locations of the subparts from the main body parts using Eq.(6). The second step is done in the contour reconstruction procedure. Because there are only six main body parts to be identified in the first step, the hypothesis space is significantly reduced. We can further improve the efficiency of the search procedure by selecting the candidates of each body parts through gating on the aspect ratios of the contour segments.

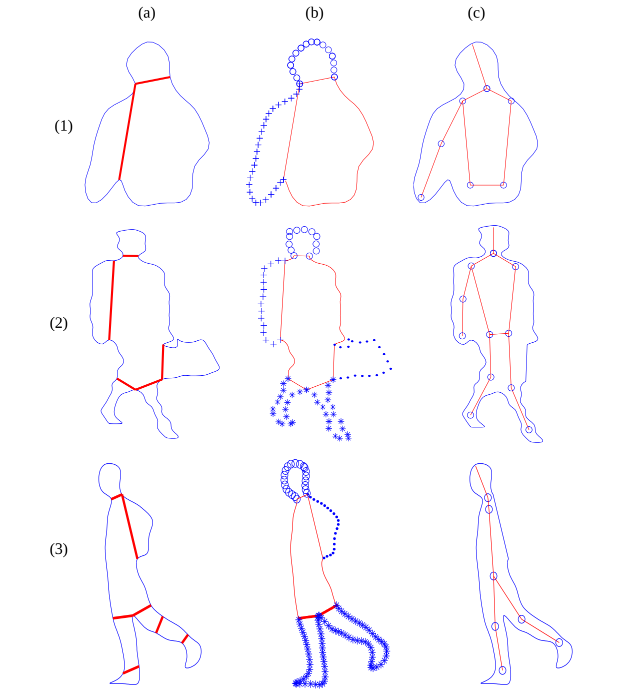


Figure 4: Body Part Identification and Localization: (a) contour partition (b) main body part identification indicated by ooo (head) — (torso) +++ (arm) *** (leg) (c) the updated locations of the identified body parts.

### 5.3 Experimental Results

Fig.4 and Fig.7 show some results of main body part identification. Here the part segmentation algorithm is based on the extraction of the convex dominant points (CDP) of the contour [14]. From the results, we can see that the proposed algorithm can identify the human body parts correctly. Even if some parts are occluded, it does not affect the identification of other parts. The algorithm can also detect the outliers simultaneously (as shown in Fig.4(b2)). The outliers may be due to the objects carried by a person, or due to the distraction from other objects. Thus, through the above body part identification procedure, we can correct certain contour extraction errors. There are also some errors with the labeling such as one shown in Fig. 4(c2). The left calf and the right calf are switched. In the future, we will incorporate the posture constraints to correct the above errors. In Fig. 4(b3), the arm is incorrectly labeled as an outlier, but as we will see in the second iteration, through the RCR algorithm we can find the missed arm correctly (as shown in Fig. 5(c3)).

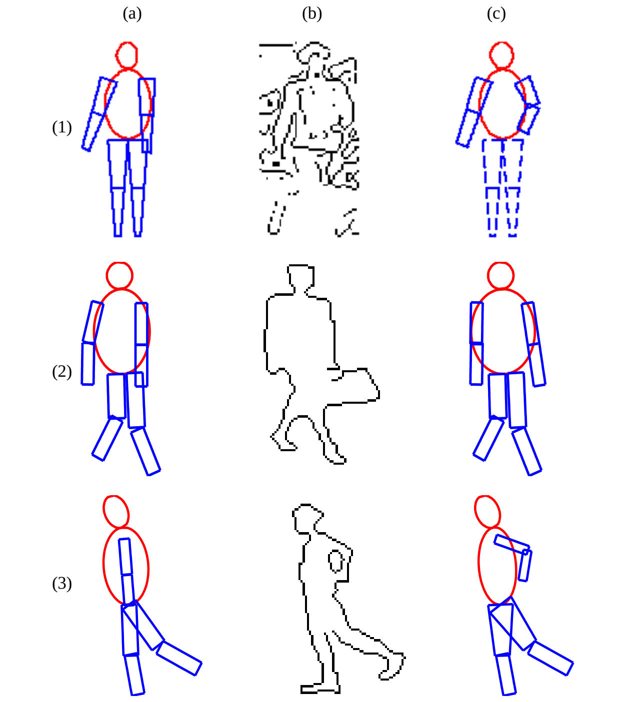


Figure 5: The second iteration: (a) the reconstructed outlines of the body parts (b) the edge images (c) the aligned body parts

## 6 Human Contour Reconstruction

The goal here is to demonstrate that using the identified human features and the human model, we can refine the shapes and locations of the identified body parts and predict the shapes and locations of the missed body parts. It includes two steps. First, integrating the identified features and the human model to refine the identified body parts. This is achieved through a $Kalman filtering$ approach. Similar frameworks have been employed in [20, 21] to locate the articulated parts sequentially. We extend the framework to include the shape and size updating besides the pose updating.

Let $C_I \sim \mathcal{N}(\hat{C}_I, \Sigma_{C_I})$ be the model matrixes estimated from the identified body parts. Let $F_I \sim \mathcal{N}(\bar{F}_I, \Sigma_{F_I})$ be the learned model matrixes. Using a Kalman Filter to integrate two estimates, we get an optimal estimation of the identified features $C_I^* \sim \mathcal{N}(\hat{C}_I^*, \Sigma_{C_I^*})$. Let $K$ be the Kalman gain factor, we have

$$K = \Sigma_{C_I}(\Sigma_{C_I} + \Sigma_{F_I})^{-1} \quad (13)$$

$$\hat{C}_I^* = \hat{C}_I + K(\bar{F}_I - \hat{C}_I) \quad (14)$$

$$\Sigma_{C_I^*} = \Sigma_{C_I} - K\Sigma_{C_I} \quad (15)$$

As demonstrated in Fig.4(c) and Fig.5(a), the integrated estimation improves the accuracy of the locations and shapes of the body parts.

The second step is to predict the shapes and locations of the unidentified body parts from the human model and the identified features. We need to estimate a feature's parameter vector $f_j = (s_j, b_j, x_j, y_j, \theta_j)$. We assume that the aspect ratio of a body part is independent of other features' aspect ratio, then the MAP estimation of $s_j$ is simply its mean value $\hat{s}_j = \bar{s}_j$.

The location of a main body part $f_j$ is estimated from the configuration vector $U$. If more than two features have been identified, then using LSM, we can estimate the transformation $T$ that projects the configuration $U$ in the model configuration space to the image coordinate system. Then we have $(x_j, y_j) = T * (u_j, v_j)$. Given the updated parameters of main body parts, we can infer the locations of the subparts through Eq.(6). Fig.4(c) shows the results of subparts localization. We know that contour partition is a very unstable procedure — a contour may be over-segmented or under-segmented. But, through this hierarchical localization procedure, we can correct the contour segmentation errors.

The length $b_j$ of a body part can be estimated from any of the identified features: $\hat{b}_j|\hat{b}_i = \bar{r}_{ji}\hat{b}_i$. If more than one features have been identified, then the MAP estimation of $b_j$ is the weighted summation derived from the Kalman Filter.

We cannot predict the orientations of the missed features, because the posture constrains are not modeled yet. This is solved in the second iteration of the RCR algorithm by aligning the predicted feature contour with the detected edge features (as shown in Fig.5(c)). Because of the cluttered background, the predicted outlines of a body part may be distracted by other objects (as shown in Fig.(8)). To avoid such situations, we need to use other cues such as stereo, motion, and the intensity pattern to constrain the search of the body parts to be within the region of similar attributes.

# 7 Human Detection

In Section 2, we developed a decision rule(as described in Eq.(2)) to perform human detection. To apply this decision rule, we need to calculate the likelihood ratio $\Lambda = P(C|w)/P(C|\bar{w})$. In the following sections, we will describe how to estimate the likelihood ratio and give the experimental results on human detection base on the derived decision rule.

## 7.1 Likelihood Ratio Estimation

To evaluate the likelihood ratio, we need to calculate two likelihoods: $P(C|w)$ and $P(C|\bar{w})$. First, we can rewrite $P(C|w)$ by conditioning on the hypotheses of the part identification:

$$P(C|w) = \sum_{H \in \mathcal{H}} P(C|H, w)P(H|w), \tag{16}$$

where $\mathcal{H}$ is the hypotheses space.

Because it is not efficient to explore all hypotheses in $\mathcal{H}$, we employ a winner-take-all strategy to approximate $P(C|w)$:

$$P(C|w) \approx P(C|H^*, w)P(H^*|w) \tag{17}$$

where $H^*$ is the optimal hypothesis selected in the part identification procedure. This approximation works well in the case of low noise and unambiguous data. Because the RCR algorithm will improve the accuracy of contour extraction iteratively. The above approximation will get more accurate accordingly.

From Eq.(10) we have

$$\Lambda \approx \frac{P(\hat{S}|H^*, w)}{P(\hat{S}|\bar{w})} \frac{P(\hat{U}|H^*, w)}{P(\hat{U}|\bar{w})} \frac{P(\hat{R}|H^*, w)}{P(\hat{\mathcal{R}}|\bar{w})} P(H^*|w) \tag{18}$$

Because $H^*$ is the best hypothesis, we have $\hat{S} \approx \bar{S}$, and $P(\hat{S}|\bar{w}) \approx P(\bar{S}|\bar{w})$. Similarly we have $P(\hat{U}|\bar{w}) \approx P(\bar{U}|\bar{w})$ and $P(\hat{R}|\bar{w}) \approx P(\bar{R}|\bar{w})$. The ratios $k_s = 1/P(\bar{S}|\bar{w})$, $k_u = 1/P(\bar{U}|\bar{w})$, and $k_r = 1/P(\bar{R}|\bar{w})$ can be considered as weights to adjust the contributions of the shape, the configuration and the size information to the final decision.

## 7.2 Experimental Results

In Fig.4 and Fig.5, we give three individual examples demonstrating that the human body parts can be identified and located correctly. In Fig.6 to 8, we describe a complete run of the RCR algorithm.

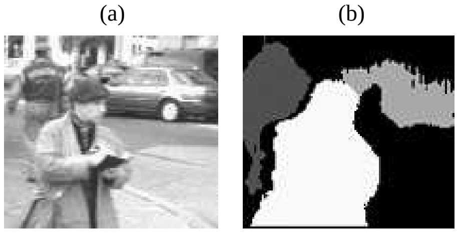


Figure 6: Stereo-based segmentation

In the first iteration, the initial outlines of objects (see Fig.6(b)) were extracted from stereo-based segmentation [24]. You can see that the contours are not

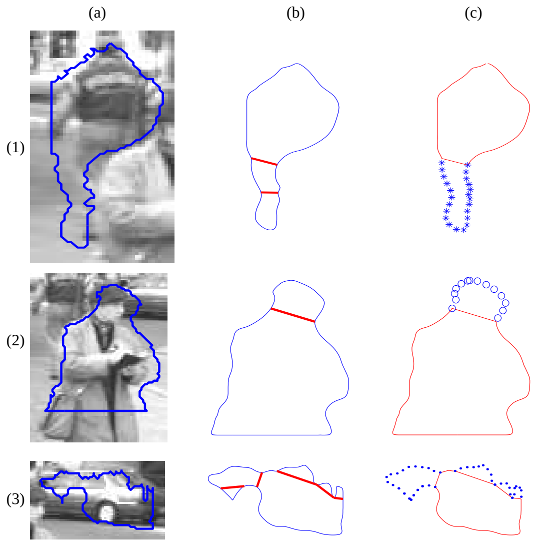


Figure 7: The first iteration of the RCR algorithm

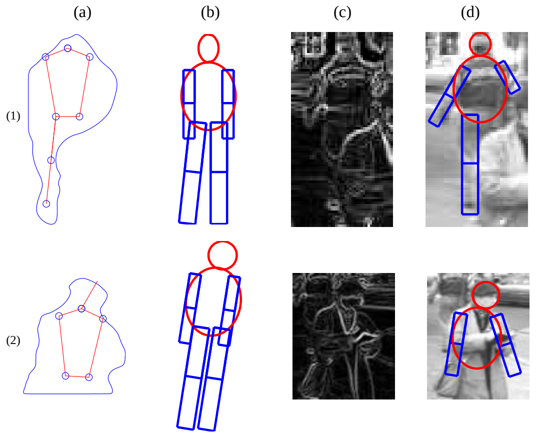


Figure 8: The second iteration of the RCR algorithm

complete and some body parts are missed due to partial occlusion or due to be out of the field of view. From the incomplete contours, the human body parts are identified as shown in Fig. 7(c). According to the decision rule given in Eq.(2), the car is correctly identified as not a person. The two people are considered as maybe humans. To solve the uncertainties, we go to the second iteration by using the reconstructed outlines of the body parts (as shown in Fig. 8(b)) as the initial contours. The orientations of the missed parts are set to be the same as that of the torso. Their actual orientations are obtained by aligning the reconstructed contours with the edge contours as shown in Fig. 8(d). With more body parts being identified, we can declare that two people are detected. In the second iteration, we can locate the missed body parts and determine if an object is a person more accurately by verifying the shapes and the locations of the body parts generated in the first iteration and solve the uncertainties with the decision in the first iteration. Usually, only two iterations are enough to achieve the correct decision.

## 8 Discussions and Future Work

In this paper, we proposed a RCR algorithm to perform and improve contour extraction, human detection, and parts identification cooperatively and iteratively. The iterative procedure helps to identify the body parts missed in the previous iteration, and to correct errors made on contour extraction and segmentation. We developed a human body model that encodes the shapes, the relative sizes, and positions of the body parts. The model is invariant to translation, rotation, and scaling, and it is deformable. We developed a Bayesian framework to perform human detection and parts identification under partial occlusion. The experimental results demonstrate the effectiveness of the proposed algorithm, but also reveal some limitations with the algorithm. First, the identification of a limb being at the left or the right side of the body may be wrong due to the absence of posture constrains. Second, the predicted outlines of a body part may be aligned to the wrong edges instead of the real contour due to the inherent ambiguity with contour features. Third, the foreshortened limbs may not be detected correctly due to the 2D-human model.

To address the above limitations, we plan (1) to include posture constrains into the human model; (2) to keep multiple hypotheses instead of just the best one to overcome the ambiguous and noisy data; (3) to extend human model to 3D and to incorporate $2\frac{1}{2}$D or $3D$ data to deal with the foreshortening problem.


## Acknowledgments

We would like to thank Steve Seitz, Robert Collins, Larry Davis, James Rehg for helpful discussions. This research is sponsored in part by PennDot.


## References


[1] D. Hogg, "Model-based Vision: a Program to See a Walking Person," *Image and Vision computing*, Vol. 1, No. 1, pp. 5-20, 1983.

[2] K. Rohr, "Towards Model-based Recognition of Human Movements in Image Sequences," *CVGIP: Image Understanding*, Vol. 59, No. 1, pp. 94-115, Jan. 1994.

[3] A. Pentland, "Recognition by Parts," *Proc. Int'l Conf. on Computer Vision*, pp. 612-620, 1987.

[4] D.A. Forsyth, M.M. Fleck, "Body Plans," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1997.

[5] I. Haritaoglu, D. Harwood, and L. Davis, "Ghost: A Human Body Part Labeling System Using Silhouettes," *Proc. Int'l Conf. on Pattern Recognition*, pp. 77-82, 1998.

[6] I. Haritaoglu, D. Harwood, L. Davis, "$W^4$—Real Time Detection and Tracking of People and their Parts," *Technical Report*, University of Maryland, Aug. 1997.

[7] I.A. Kakadiaris, D. Metsxas and R. Bajcsy, "Active Motion-Based Segmentation of Human Body Outlines," *Workshop on Articulated Motion*, 1994.

[8] M.C. Burl, M. Weber and P. Perona, "A Probabilistic Approach to Object Recognition Using Local Photometry and Global Geometry," *Proc. European Conf. on Computer Vision*, Vol. 2, pp. 628-642, 1998.

[9] T-L. Liu and D. Geiger, "Approximate Tree Matching and Shape Similarity," *Proc. Intl. Conf. on Computer Vision*, Kerkyra, Greece, 1999.

[10] P.J. Besl, N.D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, Vol. 14, No. 2, pp. 239-256, 1992.

[11] P. Mayback, *Stochastic Models, Estimation and Control*, Academic Press, New York, 1982.

[12] A. Rosenfeld, "Axial Representation of Shape," *Computer Vision, Graphics, and Image Processing*, Vol. 33, pp. 156-173, 1986.

[13] J. Ponce, "On Characterizing Ribbons and Finding Skewed Symmetries," *Computer Vision, Graphics, and Image Processing*, 52(3), pp. 328-340, 1990.

[14] M. Bennamoun, "A Contour-Based Part Segmentation Algorithm," *Proc. Int. Conf. Acoustics, Speech*, pp. 41-44, Apr. 1994.

[15] A.R. Tilley, "The Measure of Man and Woman: Human Factors in Design," H.D. Associates, NY, 1993.

[16] M.K. Leung, Y.H. Yang, "A Model Based Approach to Labeling Human Body Outlines," *Workshop on Articulated Motion*, pp. 57-62, 1994.

[17] I. Haritaoglu, et. al., "Backpack: Detection of People Carrying Objects Using Silhouettes," *Proc. Int'l Conf. on Computer Vision*, 1999.

[18] 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.

[19] S. Ju, M. Black, and Y. Yacoob, "Cardboard People: A Parameterized Model of Articulated Image Motion," *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pp. 38-44, 1996.

[20] Y. Hel-Or and M. Werman, "Constraint Fusion for Recognition and Localization of Articulated Objects", IJCV 19(1), pp. 5-28, 1996.

[21] T. Cham and J. Rehg, "Dynamic Feature Ordering for Efficient Registration", ICCV'99, pp 1084-1091, Corfu, Greece, 1999.

[22] W. Freeman, "Exploiting the Generic Viewpoint Assumption," Technical Report TR 93-15a, Mitsubishi Electric Research Labs, 1993.

[23] C. Wren, A. Azarbayejani, T. Darrel, and A. Pentland, "Pfinder: Real time Tracking of the Human Body," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, Vol. 19, No. 7, pp. 780-785, 1997.

[24] L. Zhao and C. Thorpe, "Stereo and Neural Network-Based Pedestrian Detection," *Proc. Int'l Conf. on Intelligent Transportation Systems*, Tokyo, Japan, Oct. 1999.