

# Tracing out Human Body from Image with Implementation of Coarse of Fine Strategy

Ms.Priyanka Thakare

Department of Computer Science Engineering  
J.D.College of Engineering & Management  
Nagpur-India  
e-mail: priyanka.thakare17@gmail.com

Prof.Anil Bende

Department of Computer Science Engineering  
J.D.College of Engineering & Management  
Nagpur-India  
e-mail:anilbende@gmail.com

**Abstract**— Human segmentation in photo images is a challenging and important problem that finds numerous applications ranging from album making and photo classification to image retrieval. Segmenting human from a given input image or a video is receiving increasing attention from computer vision researchers. Different techniques are classified with respect to different approaches of segmenting human i.e. Exemplar based, part based and some other methods which are using different approaches like shape priors(CRF,MRF), ACF of segmenting human from photo images. In Exemplar approach, exemplar pool is created first and then test images are matched with the exemplars or models. Whereas in part based approach human body can be recovered by assembling set of candidate parts. Both of this approach is having some drawbacks so some other methods are developed for human segmentation.

In this paper a straightforward framework to automatically recover human bodies from colour photos is proposed by employing coarse to fine strategy, first detect a coarse torso (CT) using multicue CT detection algorithm and then extract the accurate region of upper body. Then an iterative multiple oblique histogram algorithm is presented to accurately recover the lower body on human kinematics.

**Keywords**- Graph cuts, human segmentation, multicue coarse torso detection algorithm (MCTD), multiple oblique histogram (MOH).

\*\*\*\*\*

## I. INTRODUCTION

In computer vision, **image segmentation** is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.

Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. Image segmentation is the process of dividing an image into multiple parts. This is typically used to identify objects or other relevant information in digital images.

In this paper, a proposed robust framework to recover human body from photo images is used via integrating top-down body information and low-level visual cues into Graph Cuts framework. For human segmentation, there are multiple regions of body parts, such as head, torso, and legs in the image, as a result of large appearance variation. In this we divide whole-body is extracted into two subtasks, i.e., upper-body and lower-body segmentations.

A common approach that is utilized in this scheme to construct the foreground and background graphs containing the likelihood term of each node being foreground/background and the piecewise smoothness term indicating the pixels in the same region having the same labels.

Given an image  $I$  with pixels  $v$ , let  $A$  be a binary vector whose elements  $A_i$  specify assignments to pixels  $v$ . Thus, energy function is formulated as follows

$$E(A) = \epsilon \sum R_i(A_i) + \sum Q_{(i,j)} \delta(A_i, A_j)$$

The first term defines the penalties for assigning pixel  $i$  to label  $A_i$  (foreground or background), calculated by the

negative log likelihood of the foreground/background models. The second term describes the boundary properties of segmentation  $A$ , and it is large when pixels  $i$  and  $j$  are similar and close to zero when the two are quite different.

A graph can be constructed with the nodes corresponding to pixels of the image and the edges. A graph consists of two types of undirected edges, i.e. t-links (terminal links) and n-links (neighborhood links). Each pixel has two t-links connected by each terminal (i.e., source and sink), and each pair of neighboring pixels  $\{i, j\}$  in  $N$  is connected by an n-link. Then, the max-flow/min-cut algorithm will be the optimal solution.

## II. LITERATURE SURVEY

Most of the algorithms for recovering human body in a static image fall into two categories, i.e., exemplar-based and part-based approaches.

### A. Exemplar-based Approach

An exemplar pool should be constructed first and then the test images are matched with the exemplars or models. As human poses are arbitrary that's why these models not always accurately segment the human body. It is difficult to extend the method for human segmentation. It fails to cover all the situations of poses and appearance variation.

Exemplar based approach is proposed by different researchers.

Kohli et al. utilized pose-specific conditional random and stick figures for segmentation, as well as pose estimation of humans within a Bayesian framework, which has been successfully used in 3-D human pose tracking. Unlike Kumar et al., this approach does not require the laborious process of learning exemplars. Instead we use a simple articulated stickman model, which together with a CRF is used as our

shape prior. The experimental results show that this model suffices to ensure human-like segmentations as shown in fig.1. The experimental results show that the segmentation results improve considerably as we increase the amount of information in our CRF framework.

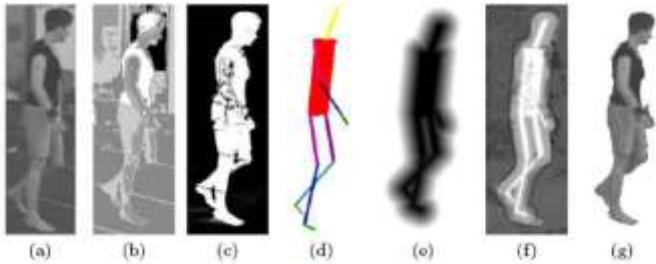


Fig.1. (a) Original image. (b) The ratios of the likelihoods of pixels being labeled foreground/background (c) The segmentation results (d) The stickman in the optimal pose .(e) The shape prior corresponding to the optimal pose of the stickman. (f) The ratio of the likelihoods of being labeled foreground/background using all the energy terms (g) The segmentation result obtained from our algorithm which is the MAP solution of the energy of the pose-specific CRF.

Lin et al. [2] proposes a hierarchical part-template matching approach for human detection and segmentation. The approach takes advantages of both local part-based and global template-based human detectors by decomposing global shape models and constructing a part-template tree to model human shapes efficiently as shown in fig.2. Edges are matched to the part-template tree efficiently to determine a reliable set of human detection hypotheses. Shape segmentations and poses are estimated automatically through synthesis of part detections. The set of detection hypotheses is optimized under a Bayesian MAP framework based on global likelihood re-evaluation and fine occlusion analysis.

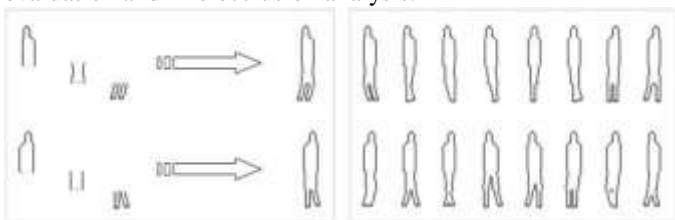


Fig.2. An illustration of the part-template tree and its construction process.

Kumar et al. represents articulated object categories using a novel layered pictorial structures model. Non articulated object categories are model using a set of exemplars as shown in fig3.



Fig.3. First and second row show multiple exemplars of the head and torso part respectively.

These models have the advantage that they can handle large intra-class shape, appearance and spatial variation. Thus we employed an efficient OBJCUT to obtain segmentation using our probabilistic framework. OBJCUT provides reliable segmentation by incorporating both:

- (i) modelled deformations, using a set of exemplars model for non-articulated objects and the LPS model for articulated objects; and

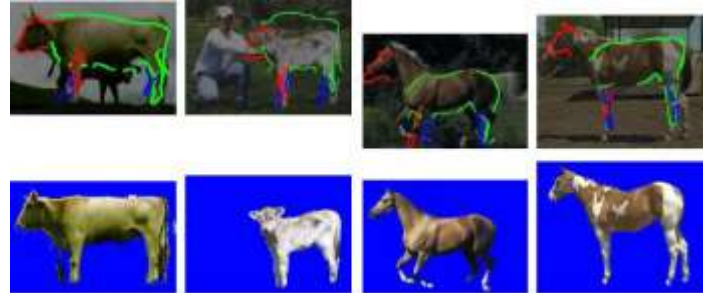


Fig.4. First row show samples of LPS model second row OBJCUT segmentation

The results for non-articulated objects are shown for two categories: bananas and oranges. The OBJCUT algorithm is also tested on two articulated object categories: cows and horses as shown in fig.4.

ACF- Adaptive contour feature is proposed for human detection and segmentation but the results for human segmentation are rough, as the labeled samples are limited.

An interactive segmentation method that incorporates local Markov Random Fields and global shape priors to estimate segmentations and pose simultaneously is also proposed. Some utilized pose-specific conditional random and stick figures for segmentation, as well as pose estimation of humans within a Bayesian framework, which has been success- fully used in 3-D human pose tracking.

In this way Kumar and Torr drew a pictorial structure and Markov random fields (MRFs) together for detecting and segmenting instances of a particular object (e.g., cows and horses) but with a limited pose variation. This method has drawback is that, these models fails to accurately segment the human body, because of arbitrary human poses and an exemplar pool cannot cover all the situations of poses and appearance variation.

### B. Part Based Approach

A different approach of segmenting human from photo images is part based approach. Part based approach is proposed by different researchers.

Mori et al. found salient half-limbs and torso by training part detectors with four main cues and Normalized Cuts. Hand-segmented limbs are used for training. However, Normalized Cuts usually do not accurately segment half-limbs and torso. The method proposed in this paper use the Normalized Cuts algorithm to group similar pixels into regions.

The method proposed in this paper use the Normalized Cuts algorithm to group similar pixels into regions. Fig. 5(e) shows segmentation with 40 regions. Many salient parts of the body pop out as single regions, such as the legs and the lower arms. In addition, we use over-segmentation, as shown in Fig. 5(f), consisting of a large number of small regions or “super pixels”, which has been shown to retain virtually all structures in real images. These segmentations dramatically reduce the complexity of later stages of analysis, e.g., from 400K pixels to 200 super pixels.

This way we detect salient upper and lower limbs from these segments. Simultaneously we detect potential head and torso positions based on exemplars to detect the torso and for detecting head some set of cues, contour, shape and focus, are used to evaluate the score of a head. The shape model of the head is simply a disk, whose scale is determined by the candidate torso. Then combine these parts into partial body configurations and prune away impossible configurations by enforcing global constraints such as relative scale and symmetry in clothing.

In the final stage a complete partial configurations by combinational search in the space of superpixels to recover full body configurations.

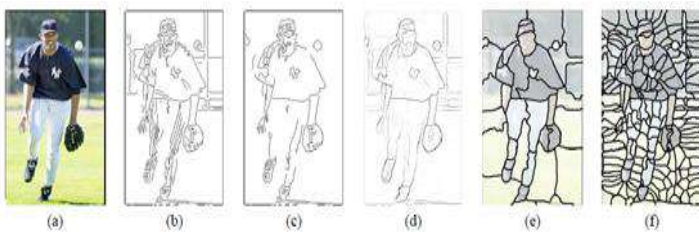


Fig.5.(a)original image(b)(c)canny edge at different scales(d) pb(probability of boundary) image (e)Normalized cut with  $k=40$  salient limbs pop out as single segments (f) “superpixel” map with 200 super-pixels

Thus for each test image, 61 images are used as the exemplars for matching. Results for this exemplar-based torso detector illustrates that this set of exemplars are unable to cope with the variation in appearance among the different players.

Ioffe et al. demonstrates probabilistic method for finding people. Since a reasonable model of a person requires at least nine segments, it is not possible to inspect every group, due to the huge combinatorial complexity. We propose two approaches to this problem. In one, the search can be pruned by using projected versions of a classifier that accepts groups corresponding to people. We describe an efficient projection algorithm for one popular classifier, and demonstrate that our approach can be used to determine whether images of real scenes contain people.

The second approach employs a probabilistic framework, so that we can draw samples of assemblies, with probabilities proportional to their likelihood, which allows us to draw human-like assemblies more often than the non-person ones. The main performance problem is in segmentation of images, but the overall results of both approaches on real images of people are encouraging.

### III PROPOSED METHODOLOGY

The proposed process is depicted in Fig . Details of each step are described below.

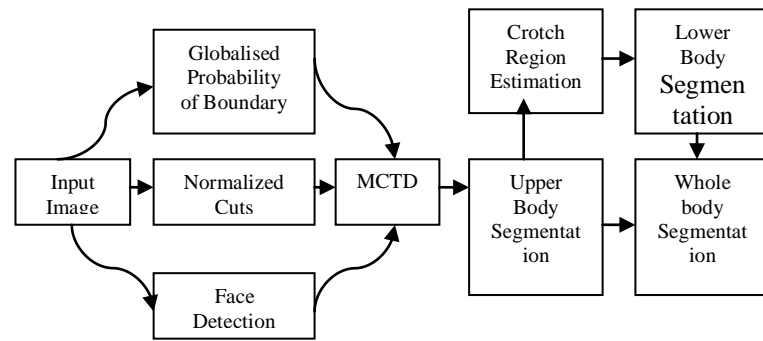


Fig.6. Block diagram of the proposed method

This paper has proposed a straightforward framework to automatically recover human bodies from color photos. It divide whole body extraction into two sub-tasks. i.e .upper body and lower bodysegmentation.In this presented MCTD algo. to detect torso and iterative MOH algo. to recover lower body segmentation. In MCTD algo. coarse torso is segmented using bounding box generated at different orientations. This bounding box is generated considering face as a priori, and then upper body is segmented using max-flow/min-cut algorithm. Coarse lower is segmented using max-flow/min- cut algo. then , iterative MOH algo. is employed to achieve fine result. But it works only on those human poses with frontal/side faces.

#### A. UPPER-BODY SEGMENTATION

Firstly use a face detection method to locate the human face from a given image. In addition to it, then, a coarse torso (CT) is detected by grouping Normalized Cuts segments. A pixel wise torso is then segmented using Graph Cuts. Torso is detected on dominant colors generated by using the k-means clustering algorithm.

##### 1) CT Detection

In this scheme, the Normalized Cuts segments are usually grouped into a torso candidate region based on the bounding box along with different orientations, where the bounding boxes are generated according to face priori. In the combining procedure, three cues are employed to select the best candidate as CT: area probability, location probability, and contour probability.

2 ) *MCTD*: Based on above three cues, the CT can be estimated with the MCTD algorithm.

By given a bounding box region  $R_i$ , first find all segments  $S_i$  that are overlapped with  $R_i$  leaving the head region. For each segment unit in  $R_i$ , we compute the area and location probability ,which can be treated as local information

$$P_{i,j} = (AP_{i,j})\lambda(LP_{i,j})1-\lambda$$

where parameter  $\lambda$  is the weighting term. The closer to the center of the bounding-box region, the more likely the segment to be a component of torso

Then, the best group for the segments corresponding to

the  $i$ th bounding-box region and its counter probability are selected by

$$CT_{i,k} = CT_{i,k}^{j^*} \{j^* = \arg \max_j (CPI_{i,k}^{j^*})\}$$

$$CPI_{i,k} = CPI_{i,k}^{j^*} \{j^* = \arg \max_j (CPI_{i,k}^{j^*})\}$$

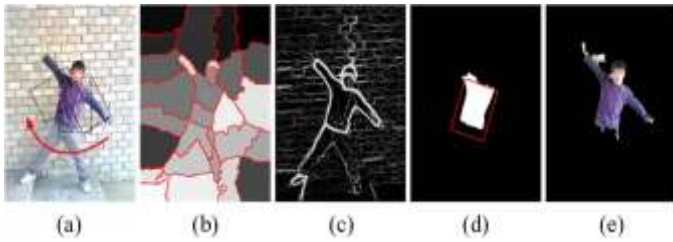


Fig.7. Upper-body segmentation. (a) Input image with bounding-box candidates.(b) Normalized Cuts segments. (c) Result of gPb. (d) CT with the bounding box. (e) Upper-body segmentation.

### B. LOWER-BODY SEGMENTATION

As the poses of legs are unpredictable, Lower-body segmentation is more challenging than upperbody segmentation. A coarse to fine scheme is developed to segment the lower part of the body.

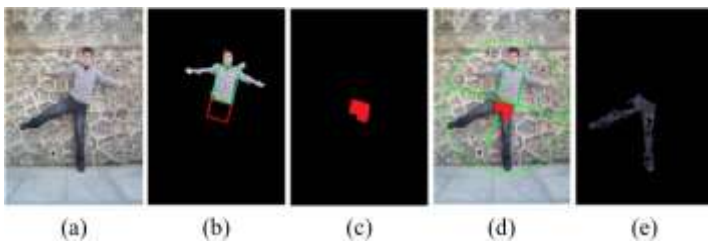


Fig.8.Coarse lower-body segmentation based on the foreground seed region. (a) Image. (b) Segmented upper body with fitted rectangle. (c) Color quantification in a initial foreground rectangle. (d) Initial hard constraints to construct a graph. (e) Coarse lower-body segmentation.

#### 1) Foreground Seed Region Initialization:

Firstly fit a torso rectangle exactly to estimate the foreground region. Estimate the torso rectangle on the segmented upper-body image, in which all the background is set in black. Torso model with four parameters (torso width, torso height, torso inclining orientation, and the neck position in the image).

#### 2) Background Distribution:

Same as the upper body, consider segmentation in a circular region containing the lower body and the pixels outside as background. The goal is to separate the lower body from the scene, so the segmented upper body can be set to be the background.

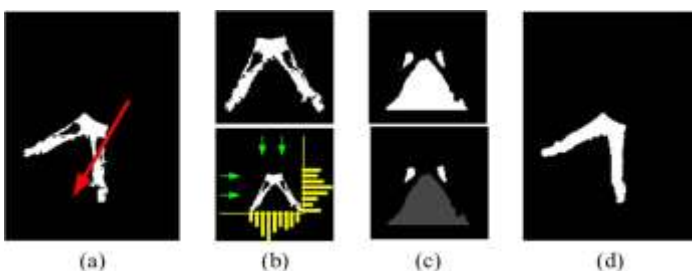


Fig.9. Lower-body segmentation based on MOH. (a) Coarse lower-body with a given orientation. (b) MOH on the projection orientation. (c) Inner blocks and their attributions. (d) Final segmentation.

#### 3) MOH-Based Lower-Body Segmentation

To obtain the fine results, an iterative MOH algorithm to refine the figure/ground distributions is implemented.

MOH is used to describe the projection information of the coarse lower body, which can be used to find the false negatives, as shown in Fig. 9(b) and (c). Each bin of MOH represents multiple cues of coarse segment results: accumulation, span, number of line segments, and boundary points of figure/ground on each projection line.

In this method, MOH is used to obtain the missed parts and judge the integrity of the lower body, so that it is used to update Graph Cuts seeds.

#### 4) Lower-Body Segmentation:

The inner blocks obtaining inner blocks employing MOH, (top), are classified into between-leg and in-leg blocks, (bottom). The gray block is regarded as the between leg block, which is selected under two evaluations. Symmetry Rule & Scale Rule.

**The symmetry** term indicates that the between-block for background seeds should be through the vertical projection line.

To constrain the scale of the between-block as a **scale** rule term.

### IV CONCLUSION

This paper has introduced a new an effective coarse to fine approach for segmentation to automatically recover human body in static photo image. The main purpose of this as follows:

- i) Proposed a segmentation-based framework for recovering human body from a static image.
- ii) To detect torso, we have MCTD.
- iii) To recover lower-body segmentation, we have introduced a robust iterative MOH algorithm.

For the future work, the algorithm to deal with variable face orientations, even in the case that the face is not possible to be detected by general face detectors.

### REFERENCES

[1] P. Kohli, J. Rihan, M. Bray, and P. Torr, "Simultaneous segmentation and pose estimation of humans using dynamic graph cut," *Int. J. Comput. Vis.*, vol. 79, no. 3, pp. 285–298, Sep. 2008.

[2] Z. Lin, L. S. Davis, D. S. Doermann, and D. DeMenthon, "Hierarchical part-template matching for human detection and segmentation," in *Proc.ICCV*, 2007, pp. 1–8.

[3] M. Kumar and P. Torr, "OBJCUT: Efficient segmentation using top-down and bottom-up cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32,no. 3, pp. 530–545, Mar. 2010

- 
- [4] G. Mori, X. Ren, A. Efros, and J. Malik, "Recovering human body configurations: Combining segmentation and recognition," in *Proc. CVPR*, 2004, pp. 326–333.
- [5] X. Ren, A. C. Berg, and J. Malik, "Recovering human body configurations using pairwise constraints between parts," in *Proc. ICCV*, 2005, pp. 824–831
- [6] G. Hua, M.-H. Yang, and Y. Wu, "Learning to estimate human pose with data driven belief propagation," in *Proc. CVPR*, 2005, pp. 747–754.
- [7] S. Ioffe and D. A. Forsyth, "Probabilistic methods for finding people," *Int.J. Comput. Vis.*, vol. 43, no. 1, pp. 45–68, Jun. 2001.
- [8] W. Gao, H. Ai, and S. Lao, "Adaptive contour features in oriented granular space for human detection and segmentation," in *Proc. CVPR*, 2009, pp. 1786–1793.
- [9] Z. Lin, L. Davis, D. Doermann, and D. DeMenthon, "An interactive approach to pose-assisted and appearance-based segmentation of human," in *Proc. ICCV*, 2007, pp. 1–8.
- [10] C. Chen and G. Fan, "Hybrid body representation for integrated pose recognition localization and segmentation," in *Proc. CVPR*, 2008, pp. 1–8.
- [11] V. Ferrari, M. Marín-Jiménez, and A. Zisserman, "2D human pose estimation in TV shows," *Statistical and Geometrical Approaches to Visual Motion Analysis*, pp. 128–147, 2009.
- [12] L. Bourdev, S. Maji, T. Brox, and J. Malik, "Detecting people using mutually consistent poselet activations," in *Proc. ECCV*, 2010, pp. 168–181.