A New Approach of Human Segmentation from Photo Images

Ashwini Magar[*], Prof.J.V.Shinde[**]

[*] Computer Department, Late G .N. Sapkal College Of Engineering, Savitribai Phule Pune University
[**] Computer Department, Late G .N .Sapkal College Of Engineering, Savitribai Phule Pune University

Abstract- This paper focuses on segmenting human from photo images. It has found several applications like album making, photo classification and image retrieval. The result can be further applied to many useful applications like part recognition which can be further applied to gesture analysis as well as in tracking. Segmenting human from photo images is still a challenge because of numerous real world factors like shading, image noise, occlusions, and background clutter and also because of great variability of shapes, poses, clothes etc. Previous works on human segmentation requires shape-matching processes. In this paper, we propose a simple method to automatically recover human bodies from photos. We use some haar cascades to detect human body that is haar cascade_upperbody and haar cascade_lowerbody which helps in performing upperbody and lower body segmentation. We need to perform CT (coarse torso) detection using MCTD algorithm for accurate upper body segmentation. Lower body is then extracted accurately using MOH based graph-cut algorithm. Experimental results show that, the proposed algorithm works well on VOC 2006 and VOC 2010 data set for segmenting person with various poses. Thus achieving high performance compared to conventional methods.

Index Terms- Graph cuts, human segmentation, Haar classifier, multicue coarse torso detection algorithm (MCTD), multiple oblique histogram (MOH).

I. INTRODUCTION

Human Segmentation from photo images is still a focus of attention in recent years as well as because of development of digital cameras intelligent processing of photos is increasingly demanded. It has found numerous applications ranging from human pose and shape estimation to photo classification and image retrieval. Along with that some high level applications on human like gait recognition can also be performed. The goal of Image segmentation is to simplify or change representation of an image into more meaningful and easier to analyze form. It is typically used to locate objects and boundaries in images. Many new approaches are proposed [1] for image segmentation. But the human segmentation from cluttered background is having less attention. Most of the pose estimation algorithm requires segmentation of human as an introductory step. For human segmentation, we have to consider multiple regions of body parts, such as head, torso, and legs in the image, as a result of large variation. Some top down cues such as shape are applied to guide segmentation [2]. But it has to face large variability of shape and appearance of a given object class. Much success has only been demonstrated in the context of object segmentation with limited pose and shape variation (e.g., pedestrian).It is impossible to have a set of model hypothesis covering all pose variations for matching. In this paper, we recover human body from still images by taking advantage of low level visual cues and some-body information into Graph Cuts framework as shown in Figure 1. It shows that we perform whole-body extraction by integrating upper-body and lower-body segmentations.

Figure 1. Human body segment ation. (a)Input image. (b)Upper-body segmentation. (c) Lower-body segmentation. (e) Final result. (f) Ground truth.

We can perform our research only on those human poses with frontal/side faces. Image segmentation can also be considered as a labelling problem. Whereas Graph cut segmentation is to construct the foreground and background graph [4] containing each node belonging to foreground or background. This way we can use graph cut segmentation to extract human object as a foreground by integrating some low level cues and thus developed coarse-to-fine scheme of human segmentation. In which we perform face detection which can then be used to estimate coarse torso using multicue coarse torso detection algorithm (MCTD) for upper body segmentation, in which image segmentation and global probability of boundary (gpb) [3] are effectively combined. Haar classifier i.e haar cascade_upperbody and haar cascade_lowerbody [17]is then applied which helps in whole body segmentation (i.e. upper body and lower body segmentation) using graph-cut to achieve a fine result. Figure 2.shows flowchart of our method. This paper focuses on segmenting the human body from photo images. These output data can be further applied to many applications. One application is to group and classify the excessive parts into legs, arms, body, and head. Face can also be label as face components such as eyes, eyebrows, nose, and mouth. Once these components are recognized, they can be used to identify the motion or gesture by the positioning of each component. For example, hands movement can be tracked for gesture and motion analysis.

Alternatively, it can be further applied to human detection and tracking in a surveillance system.

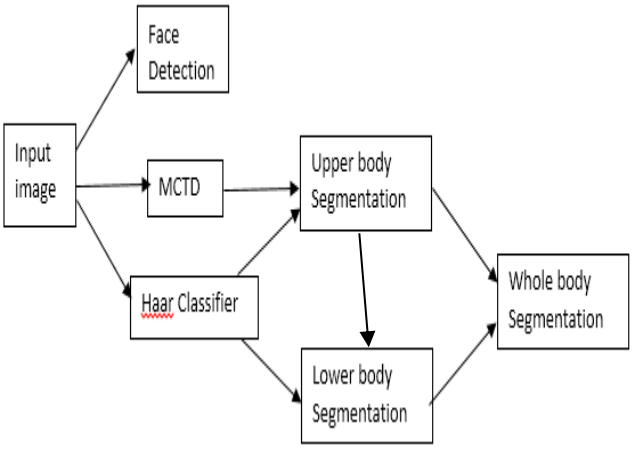

Figure. 2. Flowchart of the proposed method.

II. Literature Survey

Over the years multiple methods have been designed for human segmentation and pose estimation in images and videos. Video based methods of human segmentation are proved to be robust and effective, which can be achieved by motion cues among video frames, but it cannot be used in still images. Most of the algorithm for human segmentation from still images fall into two categories, i.e. exemplar based [2], [5] and part-based [6] approaches.

In Exemplar based method, an exemplar pool should be constructed first, and then, test images are matched with the exemplars or models. These models cannot always accurately segment the human body, because human poses are arbitrary and an exemplar pool cannot cover all the situations of poses and appearance variation. It is difficult to extend the method of [5] for human segmentation, in which Kumar and Torr represent articulated object categories using a novel layered pictorial structures model. They develop an efficient method, OBJCUT, to obtain segmentations using our probabilistic framework. We represent the shape of an object (or a part, in the case of articulated objects) using /multiple exemplars of the boundary as shown in Figure 3. Drawback of this method is that, it cannot always accurately segment body, because exemplars cannot cover all situations of poses and appearance variation.

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

Gao et al. [8] proposed an adaptive contour feature (ACF) which is effective but has rough human segmentation. Again it is also impossible to construct database of all human poses. Lin et al.[9] proposed a method that incorporates local MRF and global shape priors to iteratively estimate segmentations and pose simultaneously. Whereas, Kohlic et al. [2] use pose specific CRF for segmentation and pose estimation which has been successfully used for 3-D Human pose tracking. Unlike Kumar et al., this approach does not require the laborious process of learning exemplars. Instead they use a simple articulated stickman model, which together with an CRF is used as our shape prior. The experimental results show that this model suffices to ensure human-like segmentations.

The part-based approaches based on assembling a set of candidate parts. Ren et al. [7] developed a framework that uses arbitrary pairwise constraints between body parts. Probability of boundary (Pb)[10]can be used as part detectors and then by assembling the candidates with integer quadratic programming human body configurations are recovered. It is very hard to design a robust part detector and its performance is also limited.

Chen and Fan [11] has presented hybrid body representation. Ferrari et al. [12] use GrabCut algorithm to segment human based on head detector. In which first two stages i.e. Human detection and tracking and Foreground highlighting use a weak model of a person obtained through an upper-body detector generic over pose and appearance. This weak model only determines the approximate location and scale of the person, and roughly where the torso and head should lie. However, it knows nothing about the arms, and therefore very little about pose. Bourdev et al. [13] utilized a poselet detector to detect keypoints of human. The trained poselet masks are combined with boundary cues to detect specific poselet and finally the human is segmented. Freifed et al. [14], has proposed a contour person (CP) in which different parts are represented by different colors. Huchuan Lu et al.[15].has proposed coarse-to-fine human detection using MCTD and MOH algo which segment human from photo images. They used Normalized cut segmentation for coarse-torso detection then upper body is segmented using max-flow min-cut segmentation, then lower body is segmented using MOH algorithm. This way, they recover human body accurately with face as a priori and It will not work for non-frontal images. One more drawback is that, it requires Normalized cut segmentation which requires long processing time. It also miss some part of human to segment from photo images.

III. Implementation Details

Consider that there are N bounding boxes along different orientations relative to face , i .e .,i=1,2, 3,….,N and Ri is the area of i^{th} bounding box region and j^{th} segment overlapped with the i^{th} bounding box region R_i as $S_{i,j}$. where j=1,2,...., Li and the overlapped areas as $O_{i,j}$.

Area Probability:
The area probability indicates the j^{th} segment under the i^{th} bounding box belonging to torso or not.

Location Probability:
It specifies the likelihood of each pixel in a segment unit belonging to the given bounding box generated by haar cascade_upperbody classifier. Let W and H be width and height

of bounding box regions respectively. Thus segments not belonging to torso are removed.

Contour probability:
Maire et al. developed a gPb gives more salient edge information than local Pb is used to predict local boundary.Thus contour cue CP is the average of gPb along its boundary

B. Process Block Diagram/Algorithm

The flowchart of our algorithm is as shown in fig2.Now we will go through each module in detail. We proposed algo.in which human segmentation is performed in two tasks, i.e. upper body and lower body segmentation.

Upper Body Segmentation:

Haar classifiers are used which helps in determining Upper body and lower body region of human for whole body segmentation. A coarse torso is then detected. For coarse-torso segmentation we need a bounding box that is to be generated according to face region. So we need to perform face detection first, thus based on similarity of pixels to the torso and face regions upper body is segmented by performing max-flow/min-cut algorithm.

CT Detection:

A coarse torso (CT) is extracted via fully automatic clustering based on image segmentation. The segments are grouped into torso region based on bounding box. The bounding boxes are generated according to face region as a priori. In grouping procedure, three cues are used to select best candidate as a CT: area probability, location probability and contour probability.

Algorithm 1: *MCTD*

1: **for** $i = 1$ to N **do**
2: **for** $k = 1$ to K **do**, where k is fine tuning parameter of orientation
3: Find all segments overlapped with $R_{i,k}$
4: Remove the head region
5: **for** $j = 1$ to $L_{i,k}$ **do**
6: Sort segments by descending $P_{i,k,j}$, where $P_{i,k,j} = (AP_{i,k,j})^{\lambda}(LP_{i,k,j})^{(1-\lambda)}$
7: $CT_{i,k,\hat{j}} = CT_{i,k,\hat{j}} + S_{i,k,j}$
8: **if** $CT_{i,k,\hat{j}} > S_{\max}$, **break**
where $S_{\max} = 12S_f$ and s_f is the face area
9: **end for**
10: $CT_{i,k} = CT_{i,k,\hat{j}^*}\{\hat{j}^* = \arg\max_{\hat{j}}(CP_{i,k,\hat{j}})\}$
11: $CP_{i,k} = CP_{i,k,\hat{j}^*}\{\hat{j}^* = \arg\max_{\hat{j}}(CP_{i,k,\hat{j}})\}$
12: Update $\theta_{i,k}$ according to $CT_{i,k}$
13: **end for**
14: $Eval_{i,k} = P_{\mathrm{CT}_{i,k}} + CP_{i,k}$
15: **end for**
16: $CT = CT_{(i,k)^*}\{(i,k)^* = \arg\max_{(i,k)}(Eval_{i,k})\}$, $R_{\mathrm{CT}} = R_{(i,k)^*}\{(i,k)^* = \arg\max_{(i,k)}(Eval_{i,k})\}$

MCTD:

Based on above mentioned cues, the CT can be extracted using MCTD algorithm. For each segment unit in R_i. We compute area and location probability which is the local information then we group each segment unit into the torso according to local information in sequence. then we compute and recomputed contour probability to constrain the unlimited increase in CT. thus grouping region will converge to CT. the best group for the segments corresponding to the bounding-box region and its counter probability are selected by

$$CT_i = CT_{i,\hat{j}^*}\left\{\hat{j}^* = \arg\max_{\hat{j}}(CP_{i,\hat{j}})\right\}$$

$$CP_i = CP^*_{i,\hat{j}}\left\{\hat{j}^* = \arg\max_{\hat{j}}(CP_{i,\hat{j}})\right\}$$

Upper Body Segmentation on CT

The CT and the detected face region help in upper-body segmentation, and the t-links connecting to the upper body can be constructed by adopting kernel density estimation (KDE). Given a pixel x, the similarity between the pixel and the torso region {xi} and the face region {xj} is defined as

$$f_u(x) = \frac{1}{mnh^2}\sum_{x_i \in S_f} K\left(\frac{x - x_i}{h}\right)\sum_{x_j \in CT} K\left(\frac{x - x_j}{h}\right)$$

Where m is the number of pixels in the face region Sf , n is the number of pixels in CT, and K is the kernel function. Here, we use the Gaussian kernel with mean zero and variance h as follows.

$$K\,(x_1 - x_2 h) = \frac{1}{\sqrt{2\pi}} e^{-\frac{d^2(x_1, x_2)}{2h^2}}$$

where d(x1, x2) is the Euclidean distance between x1 and x2.Thus first we detect upper body region through haar cascade_upperbody[17] classifier. Which acts as a region of interest for upper body segmentation. Then pixels in the lines close to the waist and head belonging to the background or not can be determined according to the CT and detected face region, by comparing the similarity of the pixels to the torso and face regions, thereby much noise removed in segmentation.

Lower Body Segmentation

The coarse lower body is extracted based on max-flow/min-cut algorithm by estimating background and foreground seed using result of upper body segmentation as shown in fig.1(c).and then fine lower body is segmented from coarse lower body by using MOH algorithm which update foreground and background seed determination. So the first step is to find foreground seed and background seed for that we require to initialize a rectangle along the inclining orientation of the torso as the red rectangle. The width of bounding box is the width of the torso rectangle and height is in direct proportion to the width. Then pixels with dominant colors are set as foreground seeds based on color quantification, as shown in fig.5(c) along with that pixels which are similar to that of face region are also considered as

foreground pixels. Pixels which are outside of area of CT are considered to be background seeds.

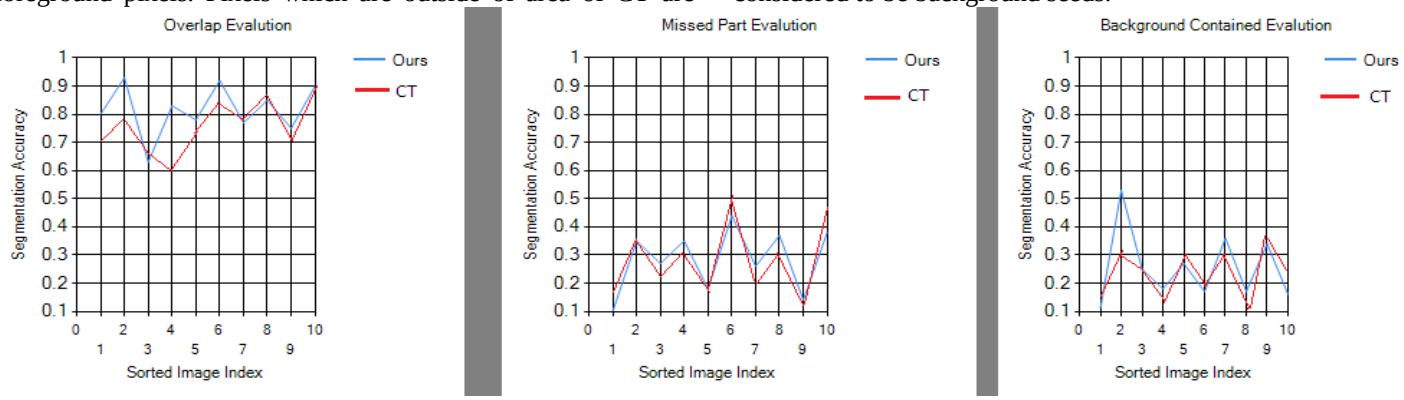

Fig.4. Evaluations of the CT detection method (red line) and ours (blue line) on our data set. (a) Overlap evaluation. (b) Missed part evaluation. (c) Background contained evaluation.

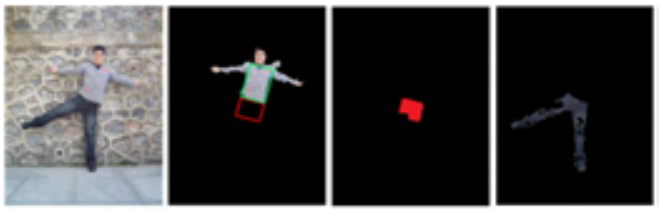

Fig.5. lower-body segmentation based on the foreground seed region. (a) Image. (b) Segmented upper body with fitted rectangle. (c) Color quantification in an initial foreground rectangle. (d) lower-body segmentation.

MOH based Lower Body Segmentation:

The inner blocks obtained is classified into in-leg and between leg .The gray block is regarded as between leg block. Coarse lower body segmentation is denoted by TS, Cs is its barycenter, N* is the inner between leg block and l is lower body orientation, which is adjusted by repeating MOH.

Algorithm 2: *Lower-body segmentation*

1: Do the first segmentation: TS
2: Initialize lower-body orientation: $l \leftarrow \hat{C}_T C_S$
3: **for** k = 1 to K **do**
4: Generate MOH on l
5: Find inner blocks
6: Select the inner between-leg block:
$N^* = N_{i^*}\{i^* = \arg\max_i(AR_i + SR_i)\}$
7: Update seeds using inner between-leg and in-leg blocks
8: Update lower-body orientation: $l \leftarrow \hat{C}_T C_i$
9: **end for**
10: Do Graph Cuts and segment lower body

IV. RESULTS

Dataset and Result set:

We have collected 197 real-world photo images with the size of 255 by 255 pixels, covering various individuals along with different poses and illumination, and cluttered backgrounds. Some samples along with the results are as in Figure 6 which indicates that human is accurately segmented from still images by our proposed algorithm.

Parameters:

In our experiments, the scale of the bounding boxes used for CT detection is set to three and four times the width and height of the detected face, respectively. The weighting terms α and β used to estimate the torso area probability are set to 0.8 and 0.5, whereas the parameter γ for location probability is set to 4, and factor λ is set to 0.4 to balance them. The width of kernel function h to calculate figure/ground distributions is 5.

For quantitative evaluation, there are three rules on the performance. The evaluations are carried out by comparing the recovered human body region with ground truth. Sinop and Grady [18] introduced a normalized overlap defined as evaluation one, which can be regarded as a global similarity between the segmentation result and the ground truth (that are manually collected). However, the rule does not consider the segmentation result how much ground truth is covered and how much background is contained. Therefore, we define another two rules to evaluate the missed ground truth and the background contained rates on the segmentation results, respectively.

Evaluation One: The similarity between the segmentation and the ground truth is defined as the proportion of pixels correctly segmented as foreground or background by comparing with ground truth binary results [18]. Where S is the set of pixels within the segmentation, and G is the set of foreground pixels in ground truth.

Evaluation Two: To evaluate how much the parts are missed, we define the second evaluation.

$$Eval1 = \frac{|S \cap G|}{|S \cup G|}$$

$$Eval2 = \frac{|G \cap \overline{S}|}{|G|}$$

$$Eval3 = \frac{|S \cap \overline{G}|}{|S|}$$

Evaluation Three: The third evaluation, which is to evaluate the background contained in the segmentation result. The evaluations on the segmentation results are shown in Figure 4. We achieve a stabler and more accurate performance than CT with less missed part and background. Table I shows, our performance is better than that of the method proposed in [19] validated on our own data set.

Table I: Performance compared with [19]

	Half-limb	Torso
Mori *et al*.[19]	81.25 %	85.50 %
Ours	96.34 %	97.65 %

Figure 6.Samples of segmentation on the VOC data set.

Ramanan [16] proposed an iterative parsing (IP) process to estimate articulated body pose. Although there is no explicit segmentation of the foreground and background, the soft segmentation results can provide the figure/ground distributions. We treat the iterative Graph Cuts algorithm employing the distributions derived from the IP method (we denote IPGC).The IPGC method poorly performs when a human subject is in cluttered background, because the IP method depends on human edge information, which is difficult to extract from the noisy scene as shown in Figure 7. In addition, much background is often segmented out along with human, and some important parts are often missed (e.g., heads or arms).So, we can say that our method is able to segment out humans in cluttered background also. Intellectual body and they select the most suitable paper for publishing after a thorough analysis of submitted paper. Selected paper get published (online and printed) in their periodicals and get indexed by number of sources.

Figure 7. Sample of segmentation by IPGC method

V. CONCLUSION

We have presented an overview of past developments in human segmentation. But recently, new technologies have made human segmentation problem popular as more convenient and affordable devices such as cameras, magnetic trackers, and computer power have become available. Many systems are based on exemplar based approach, but it fails as these models cannot always accurately segment the human body, because human poses are arbitrary and an exemplar pool cannot cover all the situations of poses and appearance variation. Related to this is the problem of how to recover from failure. A number of systems based on part based approach. Human body is recovered by assembling body parts. But problem with this approach is to develop robust part detector. Then different methods with different methodologies are developed like, Chen and Fan [11] has presented hybrid body representation, Ferrari et al. [12] utilized the GrabCut algorithm to highlight human foreground based on a head detector to reduce the search space and then detect each body part. Bourdev et al. [13] used a poselet detector to detect the keypoints of the human and train a contour person model containing shape variation, viewpoint, and rotation was defined by Freifeld et al. [14], Huchu an Lu et al[15].has proposed coarse-to-fine human detection using MCTD and MOH algo which segment human considering face as a priori. so it is still a challenging problem to obtain correct human segmentation from photo images. Our proposed method is still very simple as we used a face detector to locate head position currently. For the future work, we will relax 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.

This paper focuses on segmenting the human body from photo images. These output data can be further applied to many applications. One application is to group and classify the excessive parts into legs, arms, body, and head. Face can also be label as face components such as eyes, eyebrows, nose, and mouth. Once these components are recognized, they can be used to identify the motion or gesture by the positioning of each component. For example, hands movement as in Figure 8. can be tracked for gesture and motion analysis. Alternatively, it can be further applied to human detection and tracking in a surveillance system.

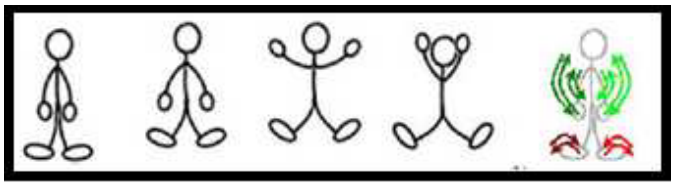

Figure 8.Application: Tracking the hand movement for motion analysis

REFERENCES

[1] W. Tao, H. Jin, and Y. Zhang, "Color image segmentation based on mean shift and normalized cuts," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*,vol. 37, no. 5, pp. 1382–1389, Oct. 2007.

[2] 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.

[3] T. Cour and J. Shi, "Recognizing objects by piecing together the segmentation puzzle," in *Proc. CVPR*, 2007, pp. 1–8.

[4] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," in *Proc. ICCV*, 2001,

[5] 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.

[6] G. Mori, X. Ren, A. Efros, and J. Malik, "Recovering human body configurations:Combining segmentation and recognition," in Proc. CVPR, 2004, pp. 326–333.

[7] X. Ren, A. C. Berg, and J. Malik, "Recovering human body configurations using pairwise constraints between parts," in Proc. ICCV, 2005,pp. 824–831.

[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] D. R. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," IEEE Trans.Pattern Anal. Mach. Intell., vol. 26, no. 5, pp. 530–549, May 2004.

[11] C. Chen and G. Fan, "Hybrid body representation for integrated pose recognition localization and segmentation," in Proc. CVPR, 2008,pp. 1–8.

[12] 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.

[13] L. Bourdev, S. Maji, T. Brox, and J. Malik, "Detecting people using mutually consistent poselet activations," in *Proc. ECCV*, 2010,pp. 168–181.

[14] O. Freifeld, A. Weiss, S. Zuffi, and M. J. Black, "Contour people: A parameterized model of 2D articulated human shape," in Proc. CVPR,2010, pp. 639–646.

[15] Huchuan Lu,Guoliang Fang, Xinqing Shao, and Xuelong Li" Segmenting Human from Photo Images Based on a Coarse-to-Fine Scheme" IEEE trans. systems ,man ,and cybernetics,Vol.42,No.3,June2012.

[16] D. Ramanan, "Learning to parse images of articulated bodies," in *Proc.NIPS*, 2006, pp. 1129–1136.

[17] Himanshu Prakash Jain, Anbumani Subramanian"Real-time Upper-body Human Pose Estimation using a Depth Camera" HP Laporatories,HPL-2010-190.

[18] A .K Sinop and L. Grady," A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm" in *Proc.ICCV*,2007,pp.1-8

[19] G. Mori, X Ren, A. Efros, and J. Malik," Recovering human body configurations: Combining segmentation and recognition," in *Proc. CVPR*,2004, pp. 326–333.

AUTHORS

First Author – Ashwini T. Magar B.E.COMP, Pune university, sulbha.magar@gmail.com.
Second Author – Prof.J.V.Shinde, M.E.COMP, Late G.N.Sapkal college of engineering.