



A survey of techniques for human segmentation from static images

Ms.Ashwini T. Magar, Prof.J.V.Shinde
Late G.N.Sapkal College of Engineering,
Computer Engineering Department, Nashik,
University of Pune, India.

Abstract: Segmenting human from a given input image or a video is receiving increasing attention from computer vision researchers. This interest is motivated by a several applications, like athletic performance analysis, surveillance, album making, man-machine interfaces, content-based image storage and retrieval. This paper review different techniques involved in segmenting human from photo images. These techniques are classified with respect 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 based 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 drawback so some other methods are developed for human segmentation. Performances are analyzed in the light of methods presented in the survey.

I. Introduction

Computer vision researchers has gained a lot of importance in last few years is understanding of human activity. Understanding human activity has applications in various fields, the most important of which is the surveillance system. Other applications include character animation for games and movies, biomechanical analysis of actions for sports and medicine etc. Human detection is a preprocessing step for activity recognition, thus segmenting human from photo images has several applications, like gesture recognition, human body tracking. This paper overview different samples of papers from broad literature on full body human detection, and presents a review and classification of various methods. Our aim is to review papers on segmenting human from a given image. The relevant literature can be divided into techniques based on different approaches. i.e Exemplar based, part based and other methods of segmenting human from still images. In Exemplar based approach [1, 2, 3] an exemplar pool should be constructed first, and then, the test images are matched with the exemplars or models. Here we review the papers which perform human segmentation using exemplar based approach (see Table1). But the problem with this approach is that in exemplar based approach the models used 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.

Table 1.Exemplar based papers

Paper	Exemplar based approach	Human feature
Kohli et al.[2007]	Dynamic Graph Cut	Pose specific CRF and stickman model.
Lin et al.[2007]	Hierarchical Part-Template Matching	Bayesian MAP framework
Kumar et al.[2005]	OBJCUT	Combines grid CDRF with object category model

In Part based approach we can recover human body configurations by assembling set of candidate parts [4, 5, 6]. Here is the list of papers that based on part based approach (see Table2). Drawback of this method is that its performance depends on individual part detector and It is very hard to design a robust part detector.

In order to overcome drawback of both exemplar and part based approach. Some different techniques are developed. Here is the list of papers [8 to 14] that used different methodologies to segment human from still images. (see Table3).

Table 2.Part based papers

Paper	Part based approach
Mori et al.[2004]	limb and torso detection using color, shape
Ren et al.[2005]	CDT graph to complete gap betn. elements
Hua et al.[2005]	DDBPMC
Ioffe et al.[2001]	Probabilistic method

Table 3.Other techniques based papers

Paper	Technique Use
Gao et al.[2009]	Active contour feature
Lin et al.[2007]	MRF & shape
Chen et al.[2008]	Hybrid method
Ferrari[2009]	Grab-cut

Bourdev[2010]	Poselet
Freifeld[2010]	Contour person model
Huchuan[2012]	Coarse-fine segmentation

II. Techniques Using Exemplar Based Approach

Exemplar based approach is proposed by different researchers. Kohli et al. [1] 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.

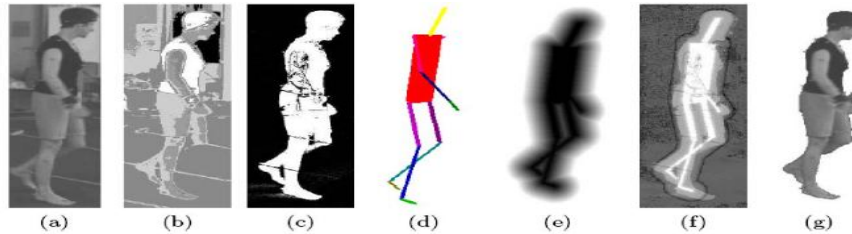
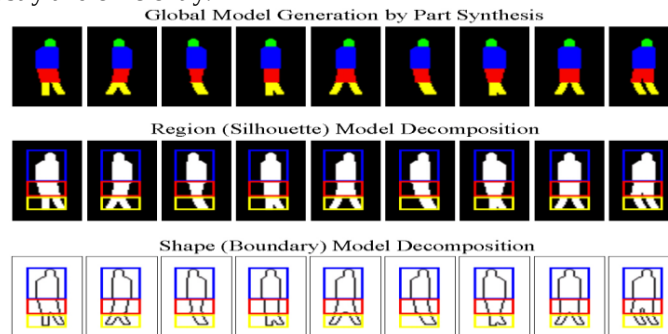


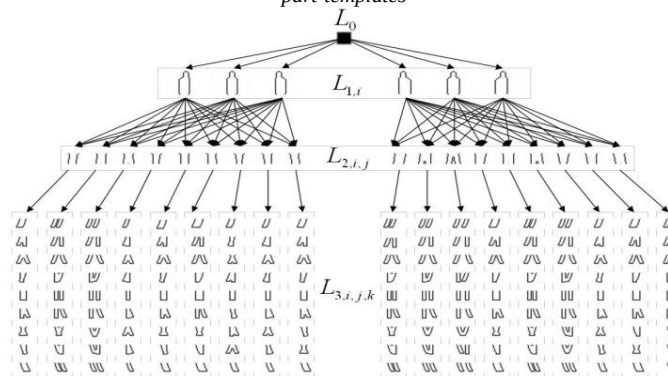
Fig.1. (a) Original image. (b) The ratios of the likelihoods of pixels being labeled foreground/background (c) The segmentation results (d) The stickman in the optimal pose. (e) The shape prior corresponding to the optimal pose of the stickman. (f) The ratio of the likelihoods of being labeled foreground/background 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.

For meeting the requirement of real-time surveillance systems, we also combined the approach with background subtraction to increase efficiency, where region information provided by foreground blobs is combined with shape information from the original image in a generalized joint likelihood model. This approach is applied to human detection and segmentation in crowded videos with and without background subtraction. The results demonstrate that the proposed part-template tree model captures the articulations of the human body, and detects humans robustly and efficiently.



(a) Generation of global shape models by part synthesis, decomposition of global silhouette and boundary models into region and shape part-templates



b) Part-template tree characterized by both shape and region information

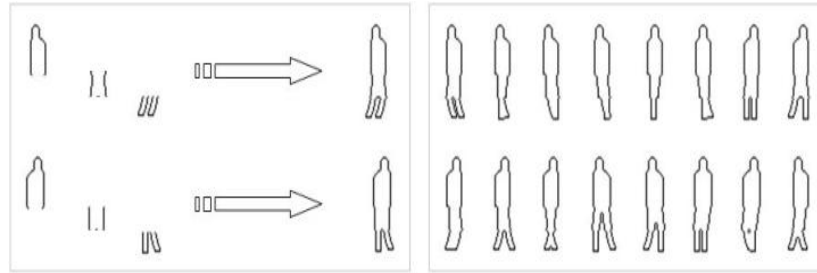


Fig.2. An illustration of the part-template tree and its construction process.

Kumar et al. [3] 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. 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 (ii) unmodelled deformations, by merging pixels surrounding the detected object into the segmentation via an st-MINCUT operation. 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.



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

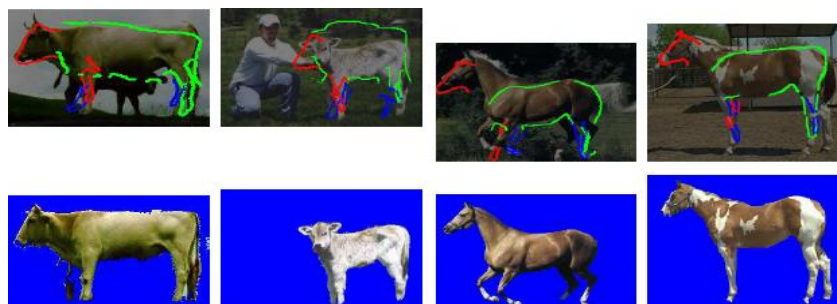


Fig.4. First row show samples of LPS model second row OBJCUT segmentation

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) with a limited pose variation. Drawback of this method is that, 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.

III. Techniques Using Part Based Approach

Part based approach is a different approach of segmenting human from photo images. Most of the research is based on this approach. Mori et al. [4] 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 as shown in fig.5. 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 “superpixels”, 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 superpixels. 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. We 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 we complete partial configurations by combinational search in the space of superpixels to recover full body configurations.

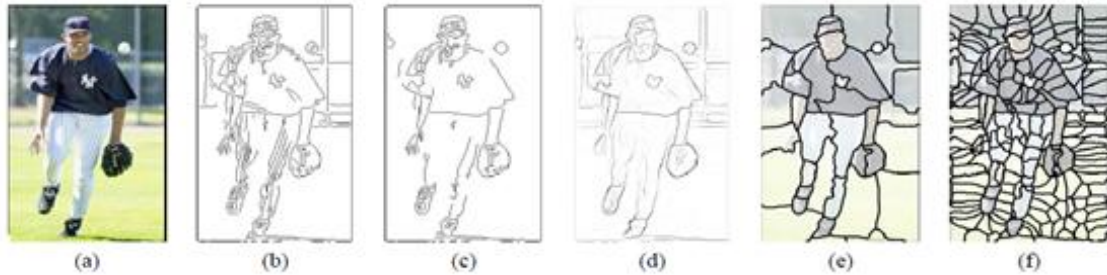


Fig.5.(a)original image(b)(c)canny edge at different scales(d) pb(probability of boundary) image (e)Normalized cut with $k=40$ salient limbs pop out as single segments (f) "superpixel" map with 200 super-pixels

Thus for each test image, we used the other 61 images 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. This way human body is recovered by combining segmentation and recognition as shown in fig.6

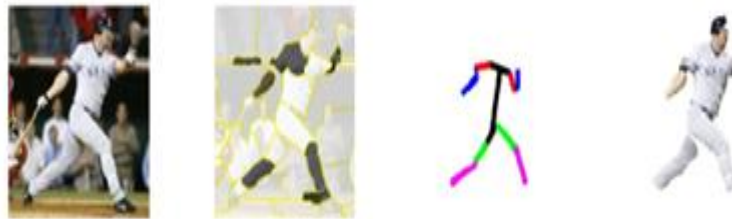


Fig.6.(a) Input image, (b) Candidate half-limbs, (c) Extracted body configurations (d) Associated segmentation.

Ren et al. [5] developed a framework which can incorporate arbitrary pair-wise constraints between body parts, such as scale compatibility, relative position, symmetry of clothing and smooth contour connections between parts. We detect candidate body parts from bottom-up using parallelism, and use various pair-wise configuration constraints to assemble them together into body configurations, as shown in fig.7.

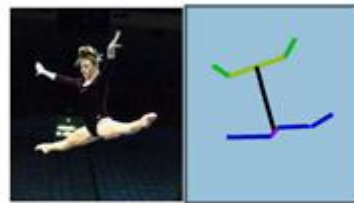


Fig.7.(a)input image(b)recovered body configuration

In which first we need to detect candidate body parts from bottom-up, and then search for the combination of the candidate parts that is most probable for human bodies. Fig. 8 shows an example of how information flows through various stages of the process: Starting with the input image in Fig. 8(a), we use the local Pb operator to compute a soft edge map in Fig. 8(b). We use Canny's hysteresis to convert the soft edge map into contours, and recursively split them into piecewise straight lines. We then use Constrained Delaunay Triangulation (CDT) to complete this scale-invariant discrete line structure into a triangulation (Fig. 8(c)). We model a body part by a pair of parallel lines and build a discriminative part detector on the basis of the CDT triangulation. We use the TRIANGLE program [20] to produce CDTs. For each pair of edges in the triangulation, we use a logistic classifier to compute its low-level saliency as a body part. The logistic classifier is trained from 15 images extracted from a skating sequence performed by Tara Lipinski with hand-labeled parts. Fig. 8(d) shows the candidate parts detected in this image. Some part candidates are detected as parallel line segments on probability of boundary (Pb) and then, the human body configurations are recovered by assembling the candidates with integer quadratic programming.

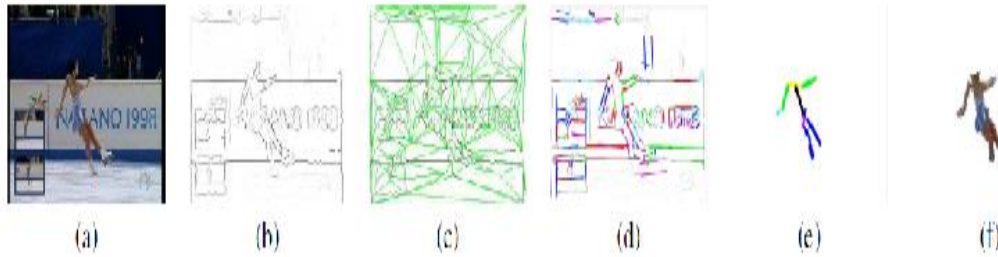


Fig.8(a)input image(b)compute an edge map(c)compute a Constrained Delaunay Triangulation(d)identify part candidates(e)Integer Quadratic Programming (f)segmentation result.

Hua *et al.* [6] proposed the data driven importance functions for body parts ,which are incorporated in the belief propagation Monte Carlo framework for efficient Bayesian inference of the human pose. Lee and Cohen [22] apply the data driven Markov Chain Monte Carlo (DDMCMC) algorithm [23] to estimate 3-D human pose from single images, in which the MCMC algorithm is utilized to traverse the pose parameter space. Nevertheless it is not clear how the detailed balance condition and convergence in the Markov chain are ensured. Most importantly, the problem of inferring 3-D body pose from single 2-D images is intrinsically ill-posed as a result of depth ambiguity. In this work, we propose a statistical formulation to infer 2-D body pose from single images. Different from the previous works, the proposed algorithm integrates the top-down and bottom-up inference with visual cues through a data driven belief propagation Monte Carlo algorithm for Bayesian reasoning. For human pose estimation, the observation models are built based on the steered edge response of the predicted body parts. The algorithm is intrinsically parallel which is in direct contrast to the sequential sampling algorithm [21] and the sequential DDMCMC approach [22].

The proposed algorithm integrates both top-down as well as bottom-up reasoning mechanism with visual cues, and carries out the inference tasks in parallel within a sound statistical framework. For concreteness, we apply the developed method to estimate pose of soccer players in single images with cluttered backgrounds. Experimental results demonstrate the potency and effectiveness of the proposed method in estimating human pose from single image as shown in fig.9.



Fig. 9. Experimental result of human pose estimation

Ioffe *et al.* [7] 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 as shown in fig.10 , 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.

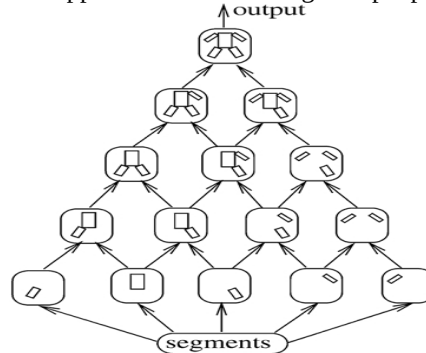


Fig.10. Pyramid of classifier

This way performance of part based human segmentation approach is depends on individual part detector and It is very hard to design a robust part detector so, its performance is also limited.

IV. Other Techniques Based Papers

Gao et al. [8] proposed an adaptive contour feature (ACF) for human segmentation as shown in fig.11. An ACF is defined as a chain of square patches called granules in Oriented Granular Space (OGS) to describe a curve. The proposed work is most related to [24, 25, 26]. However, our ACF can describe an object contour more accurately than HOG [24]. Compared with pre-defined pixel level edge-let [26], ACF is learned from the training samples, which is more invariant with pose variation. And different from [25] in which contour fragments are learned from segmented input masks, ACF is mined out in learning discriminative features for detection and is object independent.

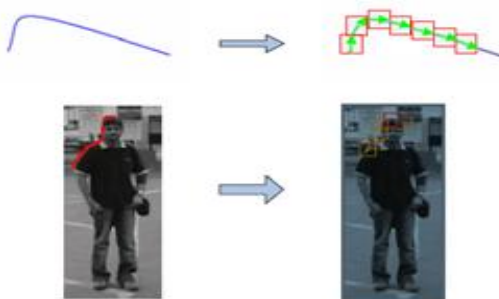


Fig.11. Describe Contour with Adaptive Contour Feature

After labeling 620 human silhouettes, they trained cascade segmentation classifiers directly with the ACF as shown in fig.12. ACF is very effective for detection as shown in fig.13, then the ACF is applied to the pixels in the ACF's effective field for segmentation. The effective field of an ACF's is defined as the pixels located in the granules of the ACF. The segmentation classifier is trained together with the cascade structure detector.



Fig.12. Samples for segmentation train.

There may be many detection responses for one human. The segmentation classifier is applied to each of these responses. The final segmentation result is achieved using a simple voting method. The results of human segmentation are rough, as the labeled samples are limited. In addition, it is almost impossible to construct a reasonable database for all the human poses.

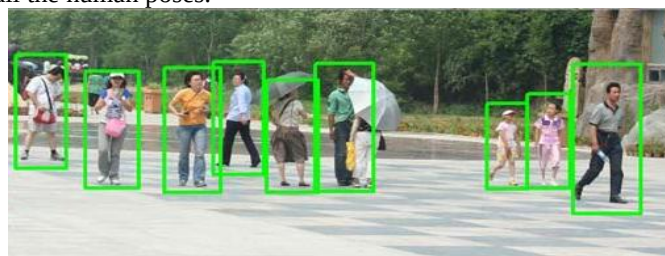


Fig.13. Detection result



Fig.14. Segmentation result

Lin et al. [9] proposed an interactive human segmentation approach. Given regions of interest provided by users, the approach iteratively estimates segmentation via a generalized EM algorithm. Specifically, it encodes both

spatial and color information in a nonparametric kernel density estimator, and incorporates local MRF constraints and global pose inferences to propagate beliefs over image space iteratively to determine a coherent segmentation. This ensures the segmented humans resemble the shapes of human poses. Additionally, a layered occlusion model and a probabilistic occlusion reasoning method are proposed to handle segmentation of multiple humans in occlusion. The approach is tested on a wide variety of images containing single or multiple occluded humans as shown in fig 15.

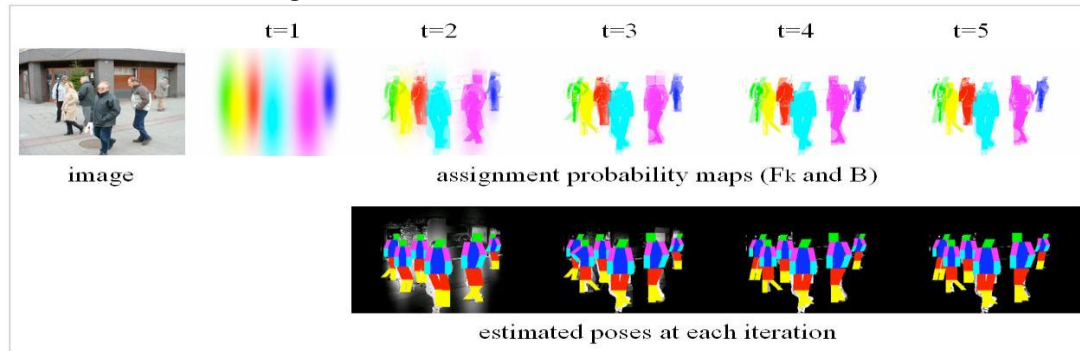


Fig.15.The process of pose-assisted segmentation for multiple occluded objects.

Chen et al. [10] has presented the hybrid body representation method integrating pose recognition, localization, and segmentation into a whole framework with restricted pose class (e.g., walking) to segment human as shown in fig.16. Recently, the idea of using segmentation to verify object hypotheses has achieved remarkable success in object detection [28, 29]. We here by propose a new hypotheses-and-test model. The method in [27] is similar in spirit to the part-level inference proposed here. This method is deeply inspired and motivated by shape representation theories in cognitive psychology where there are two prevailing theories, i.e., the structural description-based and the view-based representations integrated into hybrid body representation to support integrated pose recognition, localization, and segmentation. The method constrains pose variation in a few classes, so the application is very limited. Thus this approach is applicable to any articulated object that has well defined spatial configurations and can be decomposed into different parts that have little shape variability.

Our future research will focus on extending the proposed body representation to be a dynamic human body representation that supports video based pose recognition, localization, tracking and segmentation.



Fig.16.input image represented by watershed cells, the part based body representation, the online learned whole shape prior, and the part-whole segmentation results(from left to right).

Ferrari et al. [11] adopt a recent and successful approach where detection in each frame proceeds in three stages, followed by a final stage of transfer and integration of models across the frames. Proposed method utilized the GrabCut[32] algorithm to highlight human foreground based on a head detector to reduce the search space and then detect each body part for estimating pose. In this paper Ferrari et al. has used an approach in which first two stages i.e.1.Human detection and tracking and 2. 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. The purpose of the weak model is to progressively reduce the search space for body parts.

The next stages i.e. Single-frame parsing and spatiotemporal parsing switch to a stronger model, i.e. a pictorial structure [30, 31] describing the spatial configuration of all body parts and their appearance. In the reduced search space, this stronger model has much better chances of inferring detailed body part positions as shown in fig.17. This way they demonstrated upper body estimation on extremely challenging video material and it also covers all poses within the upper-body frontal range. Special attention is given to the arms, as they carry most of the information necessary to distinguish pose. This method supports arms folded over the torso, stretching outwards, pointing forward, etc. The numerous works defining action descriptors based on body outlines could benefit from our technique, as it provides outlines without resorting to traditional background segmentation, requiring a known and static background. Of course, further improvements are possible. For example the body

part segmentations could be improved by a further application of GrabCut[32] initialized from the current segmentations.

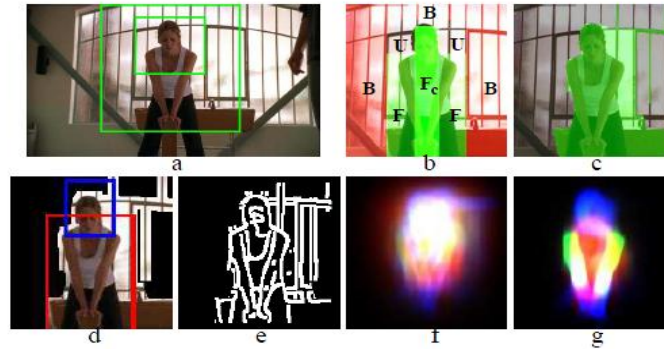


Fig. 17. Overview of the single-frame steps. 1. Upper body detection 2.Foreground highlighting: (b) subregions for initializing GrabCut. (c) foreground region output by GrabCut. 3. Parsing: (d) area to be parsed F (dilated from (c)) and (e) edges within F . (f) posterior of the part positions $p(li|I)$ after the edge-based inference. (g) posterior after the second inference, based on edges and appearance. This visualization is obtained by convolving rectangles representing body parts with their corresponding posterior.

Bourdev et al. [12] utilized a poselet detector to detect the keypoints of the human and train a figure-ground predictor of each poselet to segment the human. Bourdev and Malik [34] introduced a new notion of parts as poselets, where the key idea is to define parts that are tightly clustered both in configuration space (as might be parameterized by the locations of various joints), and in appearance space (as might be parameterized by pixel values in an image patch). Finding such parts requires extra annotation, and [34] introduced a new dataset, H3D, consisting of images of people annotated with 3D keypoints. The trained poselet masks are merged with boundary cues to locate specific poselets, and finally, the human is segmented. The main contribution is in how we use the pattern of poselet activations. Individual poselet activations are noisy, but considering the spatial context of each can provide vital disambiguating information, just as object detection can be improved by considering the detection scores of nearby objects in the scene. This can be done by training a two-layer feed-forward network with weights set using a max margin technique. The refined poselet activations are then clustered into mutually consistent hypotheses where consistency is based on empirically determined spatial keypoint distributions. Finally, bounding boxes are predicted for each person hypothesis and shape masks are aligned to edges in the image to provide segmentation.

Freifeld et al. [13] has proposed a contour person (CP) model (Fig. 19(b)) that provides a detailed 2D representation of natural body shape and captures how it varies across a population. The CP model is built from training data generated from a 3D SCAPE body model [35] capturing realistic body shape variation and non-rigid pose variation. It retains, however, the part-based representation of current 2D models as illustrated by the different colors as shown in Fig. 18(b). An articulated, part-based, model is required for pose estimation using inference methods such as belief propagation. Importantly, the CP model also captures the non-rigid deformation of the body that occurs with articulation. This allows the contour model to accurately represent a wide range of human shapes and poses.

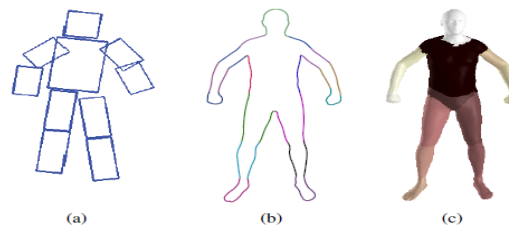


Fig.18. Contour Person Model. Most 2D body models (left),The contour person model(middle),Most 3D models(right)

This contour-person model illustrates its application in pose estimation and segmentation to do so we build on an existing state of the art person detector that uses a pictorial structures (PS) model as shown in fig2.row1. This existing technique is used to initialize our model and then both the pose and shape of the CP model are refined using a parametric form of GrabCut. Experimental results shows that CP model realistically captures a large range of real human poses in space in which it was trained. But this model does poor job representing open hands and bent wrist as shown in fig .19(g) and (f).

Huchuan et al. [14] has proposed a straightforward framework to automatically recover human bodies from color photos. They divide whole body extraction into two sub-tasks. i.e .upper body and lower body segmentation as shown in fig.21.They have 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. 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.



Fig.19. Results. Row 1: Pictorial Structures result. Row 2: CP initialization from PS (red) and CP result (green). Row 3: CP result. This method is still very simple as it uses 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.

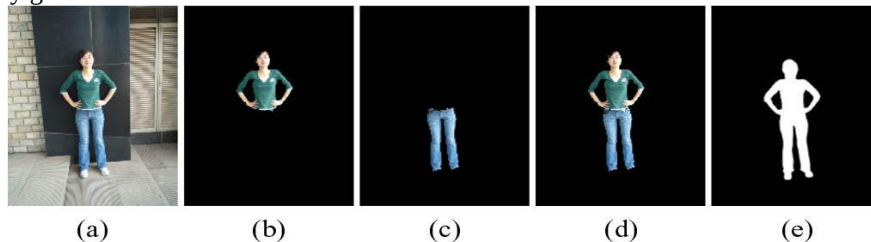


Fig.20(a)Input image(b)Upper-body segmentation(c)Lower-body segmentation(d)Final result(e)Ground truth

V. Conclusion and Future Scope

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], Huchuan 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.

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 fig.10 can be tracked for gesture and motion analysis. Alternatively, it can be further applied to human detection and tracking in a surveillance system.

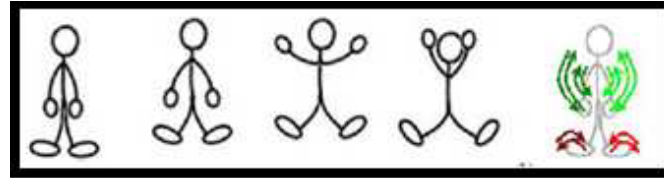


Fig.21.Application: Tracking the hand movement for motion analysis

VI. 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.
- [13] 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.
- [14] 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, June 2012.
- [15] Kohli, P., & Torr, P. (2005). Efficiently solving dynamic Markov random fields using graph cuts. In: *ICCV*.
- [16] Huang, R., Pavlovic, V., & Metaxas, D. (2004). A graphical model framework for coupling mrfs and deformable models. In: *CVPR* (Vol. II, pp. 739–746).
- [17] A. Agarwal and B. Triggs. Tracking articulated motion using a mixture of autoregressive models. In *ECCV*, pages III:54–65, 2004.
- [18] M.A. Fischler and R.A. Elschlager. The representation and matching of pictorial structures. *TC*, 22:67–92, January 1973.
- [19] D. M. Gavrila and V. Philomin. Real-Time Object Detection for SMART Vehicles. *ICCV*, 1999.
- [20] J. Shewchuk. Triangle: Engineering a 2d quality mesh generator and delaunay triangulator. In *First Workshop on Applied Computational Geometry*, pages 124–133, 1996.
- [21] S. Ioffe and D. Forsyth. Finding people by sampling. In *Proc. IEEE International Conference on Computer Vision* pages 1092–1097, 1999.
- [22] M. W. Lee and I. Cohen. Proposal maps driven MCMC for estimating human body pose in static images. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 334–341, 2004.
- [23] Z. Tu and S.-C. Zhu. Image segmentation by data-driven Markov chain Monte Carlo. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(5):657–673, 2002.
- [24] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.
- [25] V. Ferrai, T. Tuytelaars, and L. Gool. Object detection by contour segment networks. In *ECCV*, 2006.
- [26] B. Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In *Proc. 10th Int. Conf. Computer Vision*, 2005.
- [27] D. Ramanan. Learning to parse images of articulated bodies. In *Proc. of Neural Info. Proc. Systems (NIPS)*, 2006. 2.
- [28] D. Ramanan. Using segmentation to verify object hypotheses. In *Proc. of IEEE CVPR*, 2007.
- [29] B. Leibe, E. Seemann, and B. Schiele. Pedestrian detection in crowded scenes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [30] P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. *IJCV*, 61(1), 2005.
- [31] D. Ramanan, D. A. Forsyth, and A. Zisserman. Strike a pose: Tracking people by finding stylized poses. In *CVPR*, volume 1, pages 271–278, 2005.
- [32] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: interactive foreground extraction using iterated graph cuts. 23(3):309–314, 2004.
- [33] J. Sivic, M. Everingham, and A. Zisserman. Person spotting: video shot retrieval for face sets. In *International Conference on Image and Video Retrieval (CIVR 2005)*, Singapore, 2005.
- [34] Bourdev, L., Malik, J.: Poselets: Body part detectors trained using 3d human pose annotations. *ICCV* (2009)
- [35] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: Shape completion and animation of people. *SIGGRAPH*, 24(3):408–416, 2005.