A NOVEL METHOD FOR EXTRACTING STANDING HUMAN BODIES FROM SINGLE IMAGES

Dr.Mahaboob Shaik,
Assistant Professor, EED, Mjcet, Hyderabad
44maha@gmail.com

Abstract—
Segmentation of human bodies in images is a challenging task that can facilitate numerous applications, like scene understanding and activity recognition. In order to cope with the highly dimensional pose space, scene complexity, and various human appearances, the majority of existing works require computationally complex training and template matching processes. We propose a bottom-up methodology for automatic extraction of human bodies from single images, in the case of almost upright poses in cluttered environments. The position, dimensions, and color of the face are used for the localization of the human body, construction of the models for the upper and lower body according to anthropometric constraints, and estimation of the skin color. Different levels of segmentation granularity are combined to extract the pose with highest potential. The segments that belong to the human body arise through the joint estimation of the foreground and background during the body part search phases, which alleviates the need for exact shape matching.

I. Introduction:

Extraction of the human body in unconstrained still images is challenging due to several factors, including shading, image noise, occlusions, background clutter, the high degree of human body deformability, and the unrestricted positions due to in and out of the image plane rotations. Knowledge about the human body region can benefit various tasks, such as determination of the human layout, recognition of actions from static images and sign language recognition. Human body segmentation and silhouette extraction have been a common practice when videos are available in controlled environments, where background information is available, and motion can aid the segmentation through background subtraction. In static images, however, there are no such cues, and the problem of silhouette extraction is much more challenging, especially when we are considering complex cases Moreover, methodologies that are able to work at a frame level can also work for sequences of frames, and facilitate existing methods for action recognition based on silhouette features and body skeletonization.

The major contributions of this study address upright and not occluded poses.
1) We propose a novel framework for automatic segmentation of human bodies in single images.
2) We combine information gathered from different levels of image segmentation, which allows efficient and robust computations upon groups of pixels that are perceptually correlated.
3) Soft anthropometric constraints permeate the whole process and uncover body regions.
4) Without making any assumptions about the fore ground and background, except for

the assumptions that sleeves are of similar color to the torso region, and the lower part of the pants is similar to the upper part of the pants, we structure our searching and extraction algorithm based on the premise that colors in body regions appear strongly

II. RELATED WORK:

We classify approaches for human body segmentation into the following categories. The first includes *interactive* methods ([10]–[14]) that expect user input in order to discriminate the foreground and background. Interactive segmentation methods are useful for generic applications, and have the potential to produce very accurate results in complex cases. However, since they rely on low-level cues and do not employ object-specific knowledge, they often require user input to guide their process, and are inappropriate for many real-world problems, where automation is necessary. In general, this category differs from the other two, which are automatic and often task specific.

III. FACE DETECTION:

Localization of the face region in our method is performed using OpenCV's implementation of the Viola–Jones algorithm [33] that achieves both high performance and speed. The algorithm utilizes the Adaboost method on combinations of a vast pool of Haar-like features, which essentially aim in capturing the underlying structure of a human face, regardless of skin color.Since skin probability in our methodology is learned from the face region adaptively, we prefer an algorithm that is based on structural features of the face. The Viola–Jones face detector is prone to false positive detections that can lead to unnecessary activations of our algorithm and faulty skin detections. To refine the results of the algorithm, we propose using the skin detection method presented in [34], and the face detection algorithm presented in [35]. The skin detection method is based on color constancy and a multilayer perceptron neural network trained on images collected under various illumination conditions both indoor and outdoor, and containing skin colors of different ethnic groups. The face detection method is based on facial feature detection and localization using low-level image processing techniques, image segmentation, and graph-based verification of the facial structure.

IV. MULTIPLE-LEVEL IMAGE SEGMENTATION

Relying solely on independent pixels for complicated inferenceleads to propagation of errors to the high levels of image processing in complex real-world scenarios. There are several different sources of noise, such as the digital sensors that captured the image, compression, or even the complexity of the image itself and their effect is more severe at the pixel level. A common practice to alleviate the noise dwelling at the pixel level is the use of filters and algorithms that extract collective information from pixels. Moreover, groups of pixels express higher semantics. Small groups preserve detail and large groups tend to capture shape and more abstract structures better. Finally, computations based on super-pixels are more efficient and facilitate more flexible algorithms. In this study, we propose using an image segmentation method, in order to process pixels in more meaningful groups. However, there are numerous image segmentation algorithms, and the selection of an appropriate one was based on the following criteria. First, we require the algorithm to be able to preserve strong edges in the image, because they are a good indication of boundaries between semantically different regions. Second, another desirable attribute is the production of segments with relatively uniform sizes. Studies on image segmentation methods [36], [37] show that although these algorithms approach the problem in different ways, in general, they utilize low-level image cues and, thus, their results cannot guarantee compliance with the various and subjective human interpretations. Thus, we deem this step as a high-level filtering process and prefer to over-segment the image; therefore, as not to lose detail. Region size uniformity is important because it restrains the algorithm from being tricked by over-segmenting local image patches of high entropy (e.g., complex and high detailed textures) at the expense of more homogeneous regions that could be falsely merged, although they belong to semantically different objects (e.g., human hand over a wooden surface with color similar to skin).

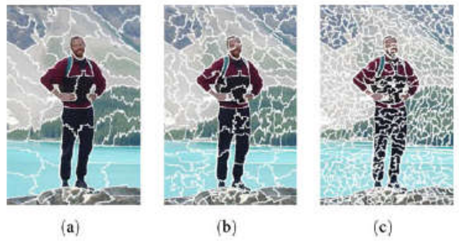

Fig. 2. Image segmentation for 100, 200, and 500 superpixels.

V. SKIN DETECTION

Among the most prominent obstacles to detecting skin regions in images and video are the skin tone variations due to illumination and ethnicity, skin-like regions and the fact that limbs often do not contain enough contextual information to discriminate them easily. In this study, we propose combining the global detection technique [39] with an appearance model created for each face, to better adapt to the corresponding human's skin color (Fig. 3). The appearance model provides strong discrimination between skin and skin-like pixels, and segmentation cues are used to create regions of uncertainty. Regions of certainty and uncertainty comprise a map that guides the GrabCut algorithm, which in turn outputs the final skin regions. False positives are eliminated using anthropometric constraints and body connectivity.

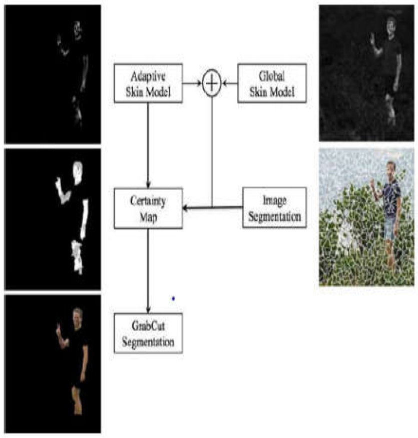

Fig. 3. Skin detection algorithm.

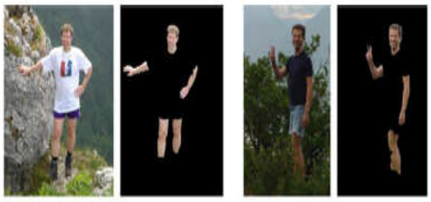

Fig: skin Detection Example

VI. UPPER BODY SEGMENTATION

In this section, we present a methodology for extraction of the whole upper human body in single images, extending [40], which dealt with the case, where the torso is almost upright and facing the camera. The only training needed is for the initial step of the process, namely the face detection and a small training set for the global skin detection process. The rest of the methodology is mostly appearance based and relies on the assumption that there is a connection between the human body parts. Processing using super-pixels instead of single pixels, which are acquired by an image segmentation algorithm, yield more accurate results and allow more efficient computations. The initial and most crucial step in our methodology is the detection of the face region, which guides the rest of the process. The information extracted in this step is significant. First, the color of the skin in a person's face can be used to match the rest of his or her visible skin areas, making the skin detection process adaptive to each person. Second, the location of the face provides a strong cue about the rough location of the torso. Here, we deal with cases, where the torso is below the face region, but without strong assumptions about in and out of plane rotations. Third, the size of the face region can further lead to the estimation of the size of body parts according to anthropometric constraints. Face detection here is primarily conducted using the Viola–Jones face detection algorithm for both frontal and side views. Since face detection is the cornerstone of our methodology, we refine the results of the aforementioned method using the face detection algorithm presented in [35].

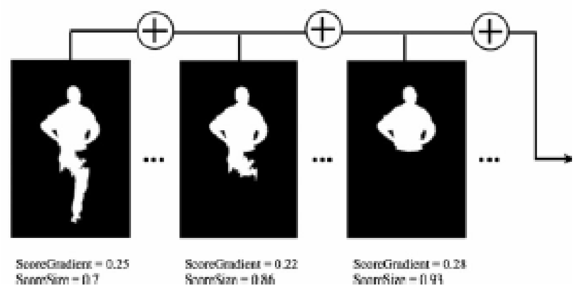

Figure: upper body segmentation

VII. LOWER BODY EXTRACTION:

The algorithm for estimating the lower body part, in order to achieve full body segmentation is very similar to the one for upper body extraction. The difference is the anchor points that initiate the leg searching process. In the case of upper body segmentation, it was the position of the face that aided the estimation of the upper body location. In the case of lower body segmentation, it is the upper body that aids the estimation of the lower body's position. More specifically, the general criterion we employ is that the upper parts of the legs should be underneath and near the torso region. Although the previously estimated UBR provides a solid starting point for the leg localization, different types of clothing like long coats, dresses, or color similarities between the clothes of the upper and lower body might make the torso region appear different (usually longer) than it should be. To better estimate the torso region, we perform a more refined torso fitting process, which does not require extensive computations, since the already estimated shape provides a very good guide.

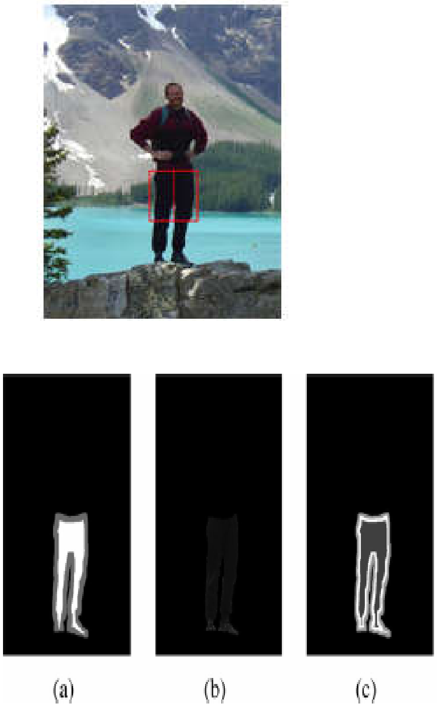

Figure: Lower body Segmentation

VIII. RESULTS:

To evaluate our algorithm, we used samples from the publicly available INRIA person dataset [41], which includes people performing everyday activities in outside environments in mostly upright position. This is a challenging dataset, since the photos are taken under various illumination conditions, in heavily cluttered environments and people appear in various types of clothing.

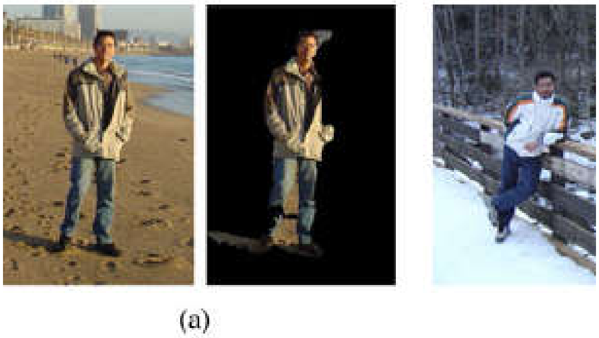

Figure: Full body Segmentation

IX. Conclusion

The first advantage of our methodology over those tested is that it can automatically localize and segment the human body. Additionally, the final results achieve very good accuracy, even in complex scenarios, and the small standard deviation shows that it is stable. The main advantages of our method are as follows. First, we combine cues from multiple levels of segmentation; therefore, to take into consideration different perceptual groupings from coarse to fine. Second, during our searching process, we try to find arbitrary salient regions that are comprised by segments that appear strongly inside the (hypothesized) foreground rectangles and weakly outside. By considering foreground and background conjunctively, we alleviate the need for exact mask fitting and dense searching, and we allow the masks to be large according to anthropometric constraints so that they may perform sufficient sampling in fewer steps. Third, we demonstrate how soft anthropometric constraints can guide and automate the process in many levels, from efficient mask creation and searching to the refinement of the probabilistic map that leads to the final mask for the body regions. Searching for the upper and lower body parts, as well as the similar process of torso fitting, however, still remain one of the most computationally expensive steps of the methodology.

REFERENCES

[1] M. Andriluka, S. Roth, and B. Schiele, “Pictorial structures revisited: People detection and articulated pose estimation,” in *Proc. IEEE Conf.Comput. Vis. Pattern Recog.*, 2009, pp. 1014–1021.

[2] M. Everingham, L. Van Gool, C. K.Williams, J.Winn, and A. Zisserman, “The pascal visual object classes (VOC) challenge,” *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.

[3] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation," in *Proc. IEEE Conf. Comput.Vis. Pattern Recog.*, 2008, pp. 1–8.

[4] M. P. Kumar, A. Zisserman, and P. H. Torr, "Efficient discriminative learning of parts-based models," in *Proc. IEEE 12th Int. Conf. Comput.Vis.*, 2009, pp. 552–559.

[5] V. Delaitre, I. Laptev, and J. Sivic, "Recognizing human actions in still images: A study of bag-of-features and part-based representations," in *Proc. IEEE Brit. Mach. Vis. Conf.*, 2010.

[6] A. Gupta, A. Kembhavi, and L. S. Davis, "Observing human-object interactions: Using spatial and functional compatibility for recognition," *IEEETrans. Pattern Anal. Mach. Intell.*, vol. 31, no. 10, pp. 1775–1789, Oct. 2009.

[7] B. Yao and L. Fei-Fei, "Grouplet: A structured image representation for recognizing human and object interactions," in *Proc. IEEE Conf. Comput.Vis. Pattern Recog.*, 2010, pp. 9–16.

[8] P. Buehler, M. Everingham, D. P. Huttenlocher, and A. Zisserman, "Long term arm and hand tracking for continuous sign language TV broadcasts," in *Proc. 19th Brit. Mach. Vis. Conf.*, 2008, pp. 1105–1114.

[9] A. Farhadi and D. Forsyth, "Aligning ASL for statistical translation using a discriminative word model," in *Proc. IEEE Comput. Soc. Conf. Comput.Vis. Pattern Recog.*, 2006, pp. 1471–1476.

[10] L. Zhao and L. S. Davis, "Iterative figure-ground discrimination," in *Proc. 17th Int. Conf. Pattern Recog.*, 2004, pp. 67–70.

[11] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.

[12] C. Rother,V.Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, Aug. 2004.

[13] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman,"Geodesic star convexity for interactive image segmentation," in *Proc.IEEE Conf. Comput. Vis. Pattern Recog.,* 2010, pp. 3129–3136.

[14] Y. Y. BoykovandM.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images," in *Proc. IEEE 8th Int.Conf. Comput. Vis.,* 2001, pp. 105–112.

[15] M. P. Kumar, P. H. S. Ton, and A. Zisserman, "Obj cut," in *Proc. IEEE Comput.Soci. Conf. Comput. Vision Pattern Recog.*, 2005, pp. 18–25.

[16] S. Li,H. Lu, and L. Zhang, "Arbitrary body segmentation in static images," *Pattern Recog.*, vol. 45, no. 9, pp. 3402–3413, 2012.

[17] L. Huang, S. Tang, Y. Zhang, S. Lian, and S. Lin, "Robust human body segmentation based on part appearance and spatial constraint," *Neurocomputing*, vol. 118, pp. 191–202, 2013.

[18] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *Int. J. Comput.Vis.*, vol. 61, no. 1, pp. 55–79, 2005.

[19] D. Ramanan, "Learning to parse images of articulated bodies," *Adv. Neur. Inf. Process. Sys.*, pp. 1129–1136, 2006.

[20] M. Eichner and V. Ferrari, "Better appearance models for pictorial structures," in *Proc. Brit. Mach. Vis. Conf.*, 2009.

[21] Y. Bo and C. C. Fowlkes, "Shape-based pedestrian parsing," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 2265–2272.

[22] Z. Hu, G. Wang, X. Lin, and H. Yan, "Recovery of upper body poses in static images based on joints detection," *Pattern Recog. Lett.*, vol. 30, no.
5, pp. 503–512, 2009.

[23] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image

segmentation," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 7–27, 2001.

[24] M. Yao and H. Lu, "Human body segmentation in a static image with multiscalesuperpixels," in *Proc. 3rd Int. Conf. Awareness Sci. Technol.*, 2011, pp. 32–35.

[25] Y. Hu, "Human body region extraction from photos," in *Proc. Mach. Vis. Appl.*, 2007, pp. 473–476.

[26] Z. Lin, L. Davis, D. Doermann, and D. DeMenthon, "An interactive approach to pose-assisted and appearance-based segmentation of humans," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

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

[28] G. Mori, X. Ren, A. A. Efros, and J. Malik, "Recovering human body configurations: Combining segmentation and recognition," in *Proc. IEEEComput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2004, pp. II-326–II-333.

[29] S. Li, H.-C. Lu, X. Ruan, and Y.-W. Chen, " Human body segmentation based on deformable models and two-scale superpixel," *Pattern Anal.Appl.*, vol. 15, no. 4, pp. 399–413, 2012.

[30] H. Lu, G. Fang, X. Shao, and X. Li, "Segmenting human from photo images based on a coarse-to-fine scheme," *IEEE Trans. Syst., Man, Cybern.B, Cybern.*, vol. 42, no. 3, pp. 889–899, Jun. 2012.

[31] S. Li, H. Lu, and X. Shao, "Human body segmentation via data-driven graph cut," *IEEE Trans. Cybern.*, vol. 44, no. 11, pp. 2099–2108, Nov. 2014.

[32] I. Rauschert and R. T. Collins, "A generative model for simultaneous estimation of human body shape and pixel-level segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 704–717.

[33] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. PatternRecog.*, 2001, pp. I-511–I-518.

[34] P. Kakumanu, S. Makrogiannis, R. Bryll, S. Panchanathan, and N. Bourbakis, "Image chromatic adaptation using ANNs for skin color adaptation,"in*Proc. 16th IEEE Int. Conf. Tools Artif.Intell. Conf.*, 2004, pp. 478–485.

[35] A. Tsitsoulis and N. Bourbakis, "A methodology for detecting faces from different views," in *Proc. IEEE 24th Int. Conf. Tools Artif.Intell.,*2012, pp. 238–245.

[36] A. Tsitsoulis and N. Bourbakis, "An LG-graph-based early evaluation of segmented images," *Meas. Sci. Technol.*, vol. 23, no. 11;

[37] H. Zhang, J. E. Fritts, and S. A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," *Comput. Vis. Image Understanding*, vol. 110, no. 2, pp. 260–280, 2008.

[38] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 2097–2104.

[39] C. O. Conaire, N. E. O'Connor, and A. F. Smeaton, "Detector adaptation by maximizing agreement between independent data sources," in *Proc.IEEE Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–6.

[40] M. C. Shin, K. I. Chang, and L. V. Tsap, "Does color space transformation make any difference on skin detection?" in *Proc. IEEE 6thWorkshop Appl.Comput. Vis.*, 2002, pp. 275–279.

[41] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Vision Pattern Recog.,Comput. Soc.Conf.*, Jun. 2005, pp. 886–893.

[42] I. EndresandD.Hoiem, “Category independent object proposals,” in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 575–588.

[43] H. Jiang, “Human pose estimation using consistent max covering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1911–1918, Sep.2011.