JOURNAL OF INFORMATION, KNOWLEDGE AND RESEARCH IN ELECTRONICS AND COMMUNICATION ENGINEERING
ISSN: 0975 – 6779| NOV 16 TO OCT 17 | VOLUME – 04, ISSUE – 02

A Methodology for Extracting Standing Human Bodies from Single Images

, Mrs.S.R.Shinde[1], Mrs.R.S.Jamdar[2]

E&TC Department, RDTC SCSCOE Dhangwadi (Pune), India[1, 2]

shindeswati151@gmail.com[1] rajshree_shinde05@rediffmail.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 com-putationally 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. The performance of our algorithm is measured using 40 images (43 persons) from the INRIA person dataset and 163 images from the "lab1" dataset, where the measured accuracies are 89.53% and 97.68%, respectively. Qualitative and quantitative experimental results demonstrate that our methodology outperforms state-of-the-art interactive and hybrid top-down/bottom-up approaches.

.

KEYWORDS: Adaptive skin detection, anthropometric constraints, human body segmentation, multilevel image segmentation.

1. Introduction

Extraction of the human body in unconstrained still mages is challenging due to several factors, including shading, image noise, occlusions, background clutter, the high degree of human body deformability, and the unrestricted posi-tions 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 ac-tions from static images , and sign language recognition. Human body segmentation and silhouette extraction have been a common practice when videos are available in con-trolled environments, where background information is avail-able, 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

For human body segmentation in static images. We decompose the problem into three sequential problems: Face detection, upper body extraction, and lower body extraction, since there is a direct pair wise correlation among them. Face detection provides a strong indication about the presence of humans in an image, greatly reduces the search space for the upper body, and pro-vides information about skin color. Face dimensions also aid in determining the dimensions of the rest of the body, according to anthropometric constraints. This information guides the search for the upper body, which in turns leads the search for the lower body. Moreover, upper body extraction provides additional in-formation about the position of the hands, the detection of which is very important for several applications. The basic units upon which calculations are performed are super pixels from multi-ple levels of image segmentation. The benefit of this approach is twofold. First, different perceptual groupings reveal more meaningful relations among pixels and a higher, however, abstract semantic representation. Second, a noise at the pixel level is suppressed and the region statistics allow for more efficient and robust computations. Instead of relying on pose estimation as an initial step or making strict pose assumptions, we enforce soft anthropometric constraints to both search a generic pose space and guide the body segmentation process. An important principle is that body regions should be comprised by segments that appear strongly inside the hypothesized body regions and weakly in the corresponding background. The general flow of the methodology can be seen in Fig. 1.

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 foreground and background, except for the assumptions that sleeves are of similar color to the torso region, and the lower part

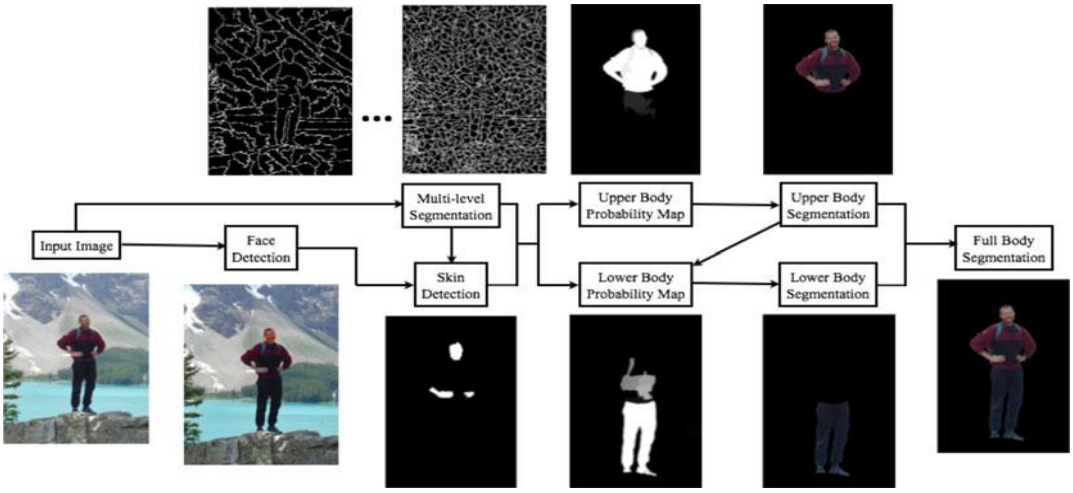

Fig. 1. Overview of the methodology. Face detection guides estimation of anthropometric constraints and appearance of skin, while image segmentation provides the image's structural blocks. The regions with the best probability of belonging to the upper body are selected and the ones that belong to the lower body follow.

2. Related Work

We classify approaches for human body segmentation into the following categories. The first includes *interactive* methods 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 au-tomation is necessary. In general, this category differs from the other two, which are automatic and often task specific.

The second category includes *top-down* approaches, which are based upon *a priori* knowledge, and use the image content to further refine an initial model. In gen-eral, human body segmentation approaches based on PS models can deal with various poses, but they rely on high-level models that might fail in complex scenarios, restricting the success of the end results. Besides, high-level inference is time consuming and, thus, these methods usually are computationally expensive.

Sampled using small masks, which is not sufficient to model the clothing in complex scenarios (no uniform clothing, cluttered background, different poses). In this methodology, we combine cues from multiple levels of segmentation. In , the human body is assumed to be inside a large mask, but due to the variability of human poses, this assumption often fails, and the sampling may lead to unrecoverable errors. In this study, we propose a more refined searching process for the torso and legs, where we try to find arbitrary salient regions in the regions that correspond to them. Salient regions are comprised of segments that appear strongly inside the hypothesized foreground and weakly in the background. By considering the 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. Pose estimation can be considered as a higher level problem compared with body segmentation, and many prefer to use a bottom-up approach to facilitate body part estimation and pose recognition .

3. Face Detection

Localization of the face region in our method is performed using OpenCV's implementation of the Viola–Jones algorithm that achieves both high performance and speed. The algo-rithm 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 detec-tions that can lead to unnecessary activations of our algorithm and faulty skin detections. To refine the results of the algo-rithm, we propose using the skin detection method presented in [34], and the face detection algorithm presented in . The skin detection method is based on color constancy and a mul-tilayer 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 lo-capitation using low-level image processing techniques, image segmentation, and graph-based verification of the facial structure.

First, the pixels that correspond to skin are detected using the method in. Then, the elliptical regions of the detected faces in the image found by the Viola–Jones algorithm are evalu-ated according to the probabilities of the inscribed pixels. More
specifically, the average skin probability of the pixels X of potential face region FR_i , for each person i, is compared with threshold $T_{G\,lo\,b\,a\,lS\,k\,in}$ (set empirically to 0.7 in our experiments). If it passes the global skin test (greater than $T_{G\,lo\,b\,a\,lS\,k\,in}$), it is further evaluated by our face detector. If the facial features are detected, then FR_i is considered to be a true positive detection.

capitation using low-level image processing techniques, image segmentation, and graph-based verification of the facial structure.

4. 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. An overview of the process can be seen in Fig. 4.

Each face region FR_j is used to construct an adaptive color model for each person's skin color. In this study, we propose using the r, g, s, I, Cr, and a channels. In more detail, $r = R/(R + G + B)$, $g = G/(R + G + B)$, and $s = (R + G + B)/3$; therefore, r and g are the normalized versions of the R and G channels, respectively, and s is used instead of b to achieve channel independence. Channels I, Cr, and a from YIQ (or NTSC), YCbCr, and Lab colorspaces, respectively, are chosen because skin color is accentuated in them. The skin color model for each person is estimated after fitting a normal distribution to each channel, using the pixels in each FR_j . The parameters that represent the model are the mean values μ_{ij} and standard deviations σ_{ij} for each FR_j and channel $j = 1 \ldots 6$ for channels r, g, s, I, Cr, and a. Each image pixel's probability of being a skin pixel is calculated separately for each channel according to a normal probability distribution with the corresponding parameters. We expect true skin pixels to have strong probability response in all of the selected channels. The skin probability for each pixel X is as follows:

$$P_{Skin}(X) = \prod_{j=1}^{6} N(X, \mu_{ij}, \sigma_{ij}). \quad (1)$$

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..

As discussed, different levels of segmentation give rise to dif-ferent perceptual pixel groupings, and each segment is described by the statistics of its color distribution. In each segmentation level, each segment is compared with the rest and its similarity image is created, depicting the

probabilistic similarity of each pixel to the segment. Similarly to the skin detection process, normal probability distributions according to the mean μ_i and standard deviation σ_i of segment S_i are estimated for each chan-nel j = 1, 2, 3 of the Lab color space, and the probability for each image pixel belonging to this probability is calculated. We estimate the final probability as the product of the probabilities. The torso is usually the most visible body part, connected to the face region and in most cases below it. Using anthro-pometric constraints, one can roughly estimate the size of the torso and its location. However, different poses and head motion make torso localization a challenging task, especially when as-sumptions about poses are relaxed. Instead of searching for the exact torso region or using complex pose estimation methods, we propose using a rough approximation of the torso mask in order to identify the most concentrated island of saliency.

5. Discussion

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 fol-lows. 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 com-prised by segments that appear strongly inside the (hypothe-sized) foreground rectangles and weakly outside. By consider-ing 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.

6. 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 inter-actions: Using spatial and functional compatibility for recognition," *IEEE Trans. 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