International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

Multi-Level Image Segmentation Model for Standing Human Body Extraction Based On Spline Regression

[1]B Chaitanya
M.Tech.,
Department of Electronics and Communications
Email: chaitanyagoud980@gmail.com

[2]T.Ravi Kumar, M.Tech, Guide, [3]**Zubair**, M.TECH, Head of Department
Department of Electronics and Communications
[1,2,3] **Dr.K.V.Subbareddy** College Of Engineering for Women, Dupadu, Kurnool

Abstract:

Digital image processing and its unmistakable quality have expanded in colossal route as of late. Computerized picture handling and related research fields clear path for the development of top of the line applications in prescription, mechanical technology, satellite picture preparing, hereditary qualities and so on. Extraction of human bodies from single pictures from particular computerized picture has achieved consideration as of late and extensive variety of research is carried on to meet the coveted outcome. A novel approach for extraction of standing human bodies has proposed in this paper where the very dimensional posture space, scene thickness, and different human appearances are dealt with in better route contrasted with customary condition of workmanship strategies. The proposed approach is arranged into five distinct advances (a) confront identification, (b) multi level division, (c) skin location, (d) abdominal area division and (e) bring down body division separately. At long last the reproduction comes about have accomplished better execution and high productivity over conventional condition of workmanship strategies.

Keywords: *Multi level segmentation, skin detection, human bodies, super pixels, bottom-up approach*

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.

In this study, we propose a bottom-up approach 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 pairwise 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 provides information about skin color. Face dimensions also aid in determining the dimensions of the rest of the body, according to anthropometric constraints. The general flow of the methodology can be seen in Fig. 1.

The major contributions of this study address upright and not occluded poses:

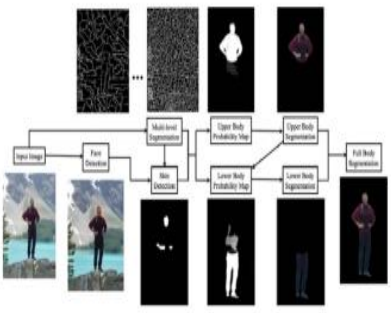
1) We propose a novel framework for automatic segmentation of human bodies in single images.

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 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 inside these regions (foreground) and weakly outside (background).

Fig.1.1. Overview of the methodology. Face detection guides estimation of anthropometric constraints and appearance of skin, while image segmentation provides the image's structural blocks.



Approach overview:

We overview the method here for the upper-body case, where there are 6 parts: head, torso, and upper/lower right/left arms. The method is also applicable to full bodies, as demonstrated.

A recent and successful approach to 2D human tracking in video has been to detect in every frame, so that tracking reduces to associating the detections. We adopt this approach where detection in each frame proceeds in three stages, followed by a final stage of transfer and integration of models across frames.

In our case, the task of pose detection is to estimate the parameters of a 2D articulated body model. These parameters are the (x, y) location of each body part, its orientation θ, and its scale. Assuming a single scale factor for the whole person, shared by all body parts, the search space has $6 \times 3 + 1 = 19$ dimensions. Even after taking into account kinematic constraints (e.g. the head must be connected to the torso), there are still a huge number of possible configurations.

Therefore, in our approach the first two stages use a weak model of a person obtained through an upper-body detector generic over pose and appearance.

The next two stages then switch to a stronger model, i.e. a pictorial structure 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.

Human detection:

We start by detecting human upper bodies in every frame, using a sliding window detection based on Histograms of Oriented Gradients, and associate detections over time. Each resulting track carves out of the total spatio-temporal volume the smaller sub volume covered by a person moving in the shot. This reduces the search space by by setting bounds on the possible (x, y) locations of the body parts and by fixing their scale, thus removing a dimension of the search space entirely.

Image segmentation:

It is the procedure of segmenting the image into different segments. It is used for Image understanding model, Robotics, Image analysis, Medical diagnosis, etc. Image segmentation means assigning a label to all pixel in the image so same labels share common visual features. Digital image having various operation like Image processing, image analysis and image understanding. In Low-level operation done by image processing and it works with pixel. Middlelevel operation done by image analysis and works with expression and description of image. High-level operation is done by image understanding and works with data symbol.

Image segmentation is the division of an image into regions or categories, which correspond to different objects or parts of objects. Every pixel in an image is allocated to one of a number of these categories. A good segmentation is typically one in which:

• pixels in the same category have similar greyscale of multivariate values and form a connected region,

• neighbouring pixels which are in different categories have dissimilar values.

For example, in the muscle fibres image, each cross-sectional fibre could be viewed as a distinct object, and a successful segmentation would form a separate group of pixels corresponding to each fibre. Similarly in the SAR image, each field could be regarded as a separate category.

Segmentation is often the critical step in image analysis: the point at which we move from considering each pixel as a unit of observation to working with objects (or parts of objects) in the image, composed of many pixels. If segmentation is done well then all other stages in image analysis are made simpler. But, as we shall see, success is often only partial when automatic segmentation algorithms are used. However, manual intervention can usually overcome these problems, and by this stage the computer should already have done most of the work.

After segmentation, methods of mathematical morphology can be used to improve the results.

In edge-based segmentation:

an edge filter is applied to the image, pixels are classified as edge or non-edge depending on the filter output, and pixels which are not separated by an edge are allocated to the same category. Fig shows the boundaries of connected regions after applying Prewitt's filter and eliminating all non-border segments containing fewer than 500 pixels. (More details will be given).

Region-based segmentation

algorithms operate iteratively by grouping together pixels which are neighbors and have similar values and splitting groups of pixels which are dissimilar in value. Fig 4.1(c) shows the boundaries produced by one such algorithm, based on the concept of watersheds, about which we will give more details.

3.3 Skin Color Segmentation:

Among various low facial features such as edge, shape, skin color and texture; skin color is prominent tool for extracting face region due to its fast processing and ease of implementation. Although color processing is advantageous but sensitive to following conditions which are discussed by Ukil Yan et al.and Nidhi Tiwari et al.:

Illumination conditions:

A change in the spectral distribution and the illumination level of light source (indoor, outdoor, highlights, shadows, color temperature of lights)

The skin color is defined by different color models like RGB, CMY, YUV, YIQ, YPbPr, YCbCr, YCgCr, YDbDr, HSV and CIE-XYZ. Comparative study and analysis of these models is done by Jose M. Chaves-Gonzalez et al. and Manuel et al.. The results of this study and analysis tell us that YCbCr, YCgCr and HSI models gives most promising results for skin segmentation and becomes most popular among others.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

Color Models

YCbCr Model:

In this color model, Y represents luminance component i.e. light intensity and Cb, Cr represents blue difference of the chromaticity component and red difference of the chromaticity component respectively.

YCgCr Model:

In the YCgCr color space, a human skin color model can be concentrated in a small region of the Cg-Cr plane. This color space includes information about green difference instead of blue difference, which can be more useful for skin color detection.

HSI Model:

In this color model, H-Hue describes the main color i.e. depth of color, S-saturation gives purity of the color and Intensity indicates the brightness of the shade. HSI color model has been used for image processing because it can separate the chromaticity from the intensity of the image.

So our interest is in combining features of these color models to get efficient face detection system.

Flow Of Proposed Work:

The flow for our proposed work is given in Fig.3.1 . The first step of dissertation is to take RGB image as input to system. This image is pre-processed by converting from RGB to appropriate color models. After this conversion, we have segmented image in two parts as skin region and non skin region by applying thresholds for each channel of model. The threshold values come from experimentation of histograms.

The 4-point and 8-point connectivity is checked on white pixels to segment face region from image. To bound face in image with rectangle, height to width ratio is applied. This ratio avoids false detections. At last, image of face with bounding box is displayed.

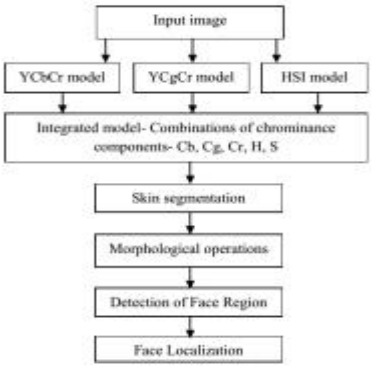


Fig.3.1. Flow of proposed work

Integration of Color Models Different combinations of chrominance components of most popular color models and their threshold values which are used for skin segmentation, shown in Table -1 .

Table -1: Integrated color models with threshold values

Combination of color models	Proposed integrated models	Threshold values for each channel
1] YCbCr, YCgCr and HSI	1] HSCbCgCr	0.4 < H < 0.8, 0.4 < S < 0.6, 82 < Cb < 160, 110 < Cg < 125, 125 < Cr < 165
2] YCgCr and HSI	1] HSCgCr 2] HCgCr	0.4 < H < 0.8, 0.4 < S < 0.6, 110 < Cg < 125, 125 < Cr < 130
3] YCbCr and YCgCr	1] CbCgCr	82 < Cb < 160, 110 < Cg < 125, 125 < Cr < 165
4] YCbCr and HSI	1] HSCbCr 2] HCbCr	0.4 < H < 0.8, 0.12 < S < 0.3, 82 < Cb < 160, 125 < Cr < 165

Fig.3.2. Implementation HSCgCr integrated color model

Discontinuity:

It means to partition an image based on immediate changes in intensity, this includes image segmentation algorithms like edge detection.

Similarity:

It means to partition an image into regions that are similar according to a set of predefined criterion. This includes image segmentation algorithms like thresholding,

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

region growing, region splitting and merging.

Edge-Based Segmentation:

An edge is a set of connected pixels that is lying on the boundary between two regions that differ in grey value. The pixels on the edge are called edge point. Edge-Based segmentation is also called as a Boundary based methods.

Parallel Edge Detection:

In parallel edge detection technique decide of whether or not a set of points are on an edge is independent. There are different types of parallel differential operators such as first difference operators and the second difference operator. The key difference between these operators is the weights allocated to each element of the mask.

Sequential Edge Detection:

In Sequential edge detection technique, the result at a point is dependent on the result of the before examined points. The act of a sequential edge detection algorithm will depend on the choice of a good initial point, and it is not easy to define termination criteria.

Region-based Segmentation:

Region based segmentation techniques split the entire image into sub regions depending on some rules.Rules like all the pixels must have the same gray level. Region-based segmentation methods attempt to group regions allowing to common image properties. Edge based methods partition an image based on rapid changes in intensity nearby edges whereas region based methods, partition an image into regions that are related according to a set of predefined criteria.

3.8.1 Region Growing:

Region growing is a procedure that group's pixels in whole image into sub regions based on predefined standard. Region Growing is used to group a collection of pixels with related properties form a region.

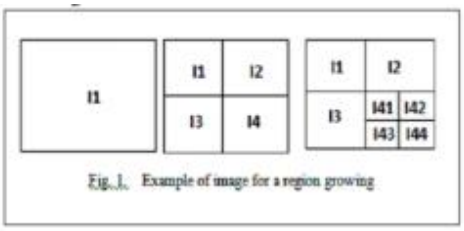


Fig. 1. Example of image for a region growing

Region Splitting and Merging:

In Region Splitting and Merging technique, the image is split into a set of arbitrary unconnected regions and merge/split the region according to the condition of the segmentation. The region split into four equal parts. Merge any adjacent regions when no more splitting is possible (see figure).

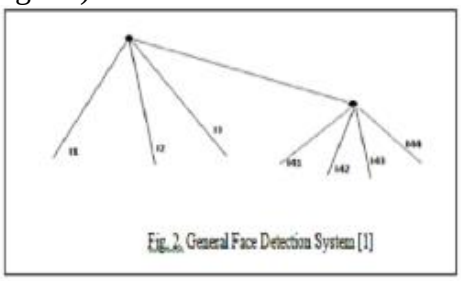


Fig. 2. General Face Detection System [1]

PROPOSED METHOD

Face Detection:

Localization of the face region in our method is performed using Open CV's implementation of the Viola–Jones algorithm 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. 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, and the face detection algorithm presented in. The skin detection method is based on color constancy and a multilayer perception 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.

After fitting an ellipse in the face region, we are able to define the fundamental unit with respect to which locations and sizes of human body parts are estimated, according to anthropometric constraints. This unit is referred to as palm length (PL), because the major axis of the ellipse is almost the same size as the distance from the base of the palm to the tip of the middle finger. Thus, our anthropometric model is adaptive for each person and invariant to scale.

Multiple-Level Image Segmentation:

Relying solely on independent pixels for complicated inference leads 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.

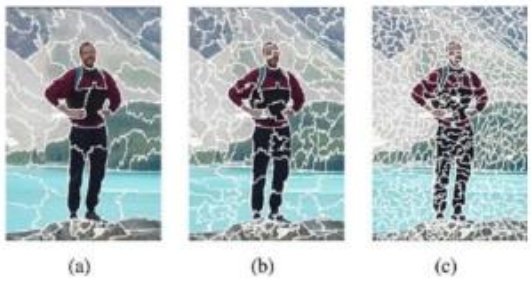


Fig. 5.1. Image segmentation for 100, 200, and 500 super pixels.

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 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 oversegment 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).

More importantly, we propose using multiple levels of segmentation, in order to alleviate the need for selecting an appropriate number for the regions to be created and combine information emanating from different perceptual groupings of pixels. Although our framework can accept any number of segmentation levels, we find that two segmentation levels of 100 and 200 segments provide accurate results. For the skin detection algorithm, a finer segmentation of 500 super pixels is used, because it manages to discriminate better between adjacent skin and skin-like regions, and recover skin segments that are often smaller compared with the rest image regions.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

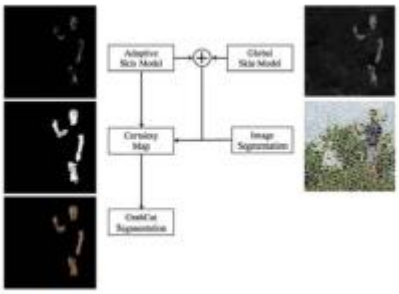


Fig. 5.2. Skin detection algorithm

5.3 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 with an appearance model created for each face, to better adapt to the corresponding human's skin color (Fig. 5.2). 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 Grab Cut 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. 5.3.

Each face region FR 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 colors paces, 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 . The parameters that represent the model are the mean values μij and standard deviations σ ij for each FR and channel j = 1 ... 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:

$$P_{Skin_i}(X) = \prod_{j=1}^{6} \mathcal{N}(X, \mu_{ij}, \sigma_{ij})$$

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

Fig. 5.3. Skin detection examples

The adaptive model in general focuses on achieving a high score of true positive cases. However, most of the time it is too "strict" and suppresses the values of many skin and skin-like pixels that deviate from the true values according to the derived probability distribution. At this point, we find that an influence of the skin global detection algorithm is beneficial because it aids in recovering the uncertain areas. Another reason we choose to extend the skin detection process is that relying solely on an appropriate color space to detect skin pixels is often not sufficient for real-world applications. The two proposals are combined through weighted averaging (with a weight of 0.25 for the global model, and 0.75 for the adaptive model). The finest level

of image segmentation is used at this point to characterize segments as certain and probable background and foreground. For the certain foreground regions, however, only the pixels with sufficiently high probability in the adaptive model are used as seeds; therefore, as to control their strong influence. In order to characterize a region as probable background or foreground, its mean probability of the combined probability must be above a certain threshold (empirically set to 0.2 and 0.3, respectively). Examples can be seen in Fig. 5.5.

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.

With respect to clothes, the size of face's ellipse guides the construction of rectangular masks for the foreground using anthropometric constraints. Our basic assumption is that a good foreground mask should contain regions that appear mostly inside the mask and not outside (background). In other words, we try to identify "islands of saliency," in the aforementioned sense. As opposed to approaches based on pose estimation, we employ simple heuristics to conduct a fast and rough torso pose estimation and guide the segmentation process.

The torso is usually the most visible body part, connected to the face region and in most cases below it. Using anthropometric 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 assumptions 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. This criterion allows for fast inference about the torso's size and location, while relieving the need for the complex task of explicit torso estimation, without sacrificing accuracy.

As discussed, different levels of segmentation give rise to different 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

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

detection process, normal probability distributions according to the mean μi and standard deviation σi of segment Si are estimated for each channel 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

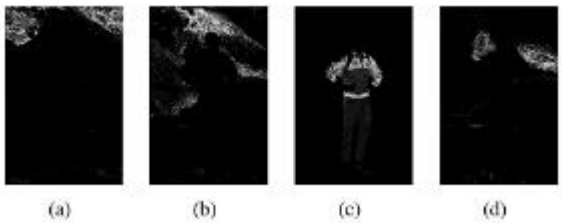


Fig. 5.4.. Example of similarity images for random segments.

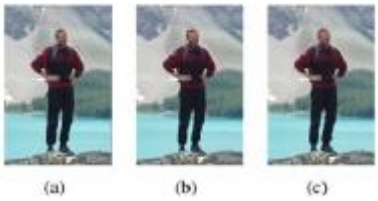


Fig. 5.5. Masks used for torso localization.

5.4.1 In Each Channel Separately:

Example similarity images are shown in Fig. 5. The resulting image that depicts the probability segment Si that is the same color as the rest of the segments is referred to as the similarity image. Similarity images are gathered for all of the different segmentation levels l. Here, we use two segmentation levels in this stage of 100 and 200 super pixels, because they provide a good tradeoff between perceptual grouping and computational complexity

$$P_{SimIm_{li}}(X) = \prod_{j=1}^{3} \mathcal{N}(X, \mu_{ij}, \sigma_{ij})$$

$$P_{SimIm_{li}}(X) = \prod_{j=1}^{3} \mathcal{N}(X, \mu_{ij}, \sigma_{ij})$$

(2)

Sequentially, a searching phase takes place, where a loose torso mask is used for sampling and rating of regions according to their probability of belonging to the torso. Since we assume that sleeves are more similar to the torso colors than the background, this process combined with skin detection actually leads to upper body probability estimation. The mask is used for sufficient sampling instead of torso fitting; therefore, it is estimated as a large square with sides of 2.5PL, with the top most side centered with respect to the face's center. In order to relax the assumptions about the position and pose of the torso, the mask is rotated by 30 ◦ left and right of its initial position (0 ◦) (see in Fig. 5.5). By using a large square mask and allowing this degree of freedom, we manage to sample a large area of potential torso locations. By constraining its size according to anthropometric constraints, we make the foreground/background hypotheses more meaningful.

During the search process, the mask is applied to each similarity image and its corresponding segment is scored. Let Torso Mask be a binary image, where pixels are set to 1 (or "on") inside the square mask and 0 (or "off") outside so that Simile ∩ Torso Mask selects the probabilities of the similarity image that appear inside the mask. Index t = 1, 2, 3 corresponds to a torso mask at angle −30, 0, or 30. Thus, (3) and (4) rate each segment's potential of belonging to the foreground and background, respectively, and (5) combines the two potentials in the form of a ratio as follows:

$$P_{FG}(S_{lli}) = \sum^{|S_{lli}|} \text{SimIm}_{li} \cap \text{TorsoMask}_t \quad (3)$$

$$P_{BG}(S_{lli}) = \sum^{|S_{lli}|} \text{SimIm}_{li} \cap \overline{\text{TorsoMask}_t} \quad (4)$$

$$\text{TorsoScore}(S_{lli}) = \frac{P_{FG}(S_{lli})}{P_{BG}(S_{lli}) + \epsilon}. \quad (5)$$

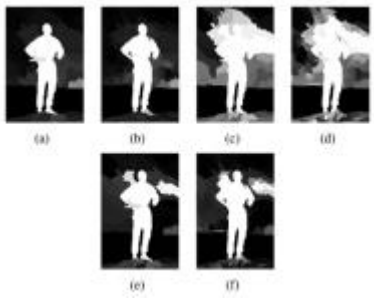


Fig.5.6. Segments with potential of belonging to torso. (a), (b) For segmentation level 1 and 2 and torso mask at 0 ◦. (c), (d) For segmentation level 1 and 2 and torso

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

mask at 30 ◦. (e) (f) For segmentation level 1 and 2 and torso mask at −30 ◦.

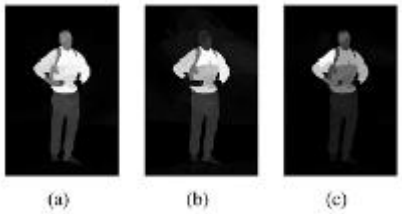


Fig. 5.7. Aggregation of torso potentials shown in Fig. 7, for torso masks at 0 ◦, 30 ◦, and −30 ◦

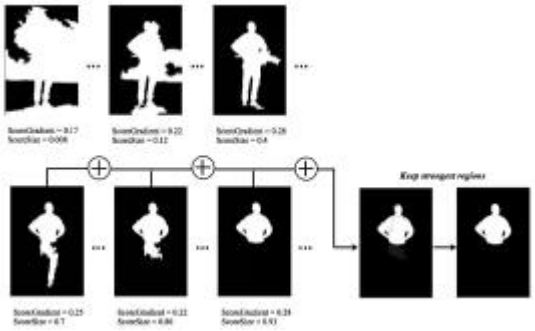


Fig. 5.8. Thresholding of the aggregated potential torso images and final upper body mask. Note that the masks in the top row are discarded.

we can achieve accurate and robust results without imposing computational strain.

The obvious step is to threshold the aggregated potential torso images in order to retrieve the upper body mask. In most cases, hands or arms' skin is not sampled enough during the torso searching process, especially in the cases, where arms are outstretched. Thus, we use the skin masks estimated during the skin detection process, which are more accurate than in the case they were retrieved during this process, since they were calculated using the face's skin color, in a color space more appropriate for skin and segments created at a finer level of segmentation. These segments are superimposed on the aggregated potential torso images and receive the highest potential (1, since the potentials are normalized).

Instead of using a simple or even adaptive thresholding, we use a multiple level thresholding to recover the regions with strong potential according to the method described, but at the same time comply with the following criteria:

1) they form a region size close to the expected torso size (actually bigger in order to allow for the case, where arms are outstretched), and

2) the outer perimeter of this region overlaps with sufficiently high gradients.

The distance of the selected region at threshold t (Region) to the expected upper body size (Exp Upper Body Size) is calculated as follows:

$$\text{ScoreSize} = \frac{e^{\frac{-|\text{Region}_t - \text{ExpUpperBodySize}|}{\text{ExpUpperBodySize}}}}{e^{\frac{-|\text{Region}_t - \text{ExpUpperBodySize}|}{\text{ExpUpperBodySize}}}} \quad (6)$$

where Exp Upper Body Size = 11 × PL2 . The score for the second criterion is calculated by averaging the gradient image (Grad Im) responses for the pixels that belong to the perimeter (Region) of Region as

$$\text{ScoreGrad} = \frac{1}{|PRegion_t|}\frac{1}{|PRegion_t|} \quad (7)$$

Thresholding starts with zero and becomes increasingly stricter at small steps (0.02). In each thresholding level, the largest connected component is rated, and the masks with Score Grad > 0.05 and Score Size > 0.6 are accumulated to a refined potential image (see in Fig. 5.8). Incorporation of this a priori knowledge to the thresholding process aids the accentuation of the true upper body regions (UBR). Accumulation of surviving masks starts when Score Size > 0.6 and resulting masks after this point will keep getting closer monotonically to the expected region size. Accumulation ends when Score Size drops below 0.6. The rationale behind this process is to both restrict and define the thresholding range and focus the interest to segments with high potential of forming the upper body segment. The aggregate mask (Aggregate Mask) can now be processed easily and produce more meaningful results. Specifically, we set a final threshold, which

allows only regions that have survived more than 20% of the accumulation process in the final mask for the UBR. This process is performed for every initial torso hypothesis; therefore, in the end, there are three corresponding aggregate masks, out of which the one that overlaps the most with the initial torso mask and obtains the highest aggregation score is selected. The aggregation score shows how many times each pixel has appeared in the accumulation process, implicitly implying its potential of belonging to the true upper body segment.

Refinement:

In many cases, the extracted upper body mask is very accurate and can be used as a final result. However, we choose to add an extra refinement step to cope with probable segmentation errors and pixels that manage to survive the multiple thresholding process. One idea that we use here is to give the upper body mask as input to an interactive foreground/background algorithm that requires "seeds" corresponding to the foreground and background. Grow Cut and Grab Cut are used for experiments.

Grow Cut expects the RGB image as input and a map denoting the seeds for background, foreground, and uncertain pixels, whereas Grab Cut can operate on a more refined map containing the certain foreground, certain background, probable foreground, and probable background regions. In order to construct these maps, we employ morphological operations on the upper body mask, with adaptive square structural elements (SEs) according to anthropometric constraints. For GrowCut, the uncertain region is constructed by dilating the upper body mask with a SE with sides equal to PL/6, the face's ellipse with a SE with sides equal to PL/10 and the skin regions with a SE with sides equal to PL/12. Possible holes between the face and torso region are also filled. The certain foreground is similarly constructed with erosion instead of dilation, where the sides of the SEs are now PL/4, PL/4, and PL/10, respectively. The rest of the map is classified as background. For the Grab Cut algorithm, the possible background ground is constructed by dilating the upper body mask, the face's ellipse and skin masks using SEs with sides PL/10, PL/2, and PL/12, the probable foreground is constructed by eroding the masks with SEs with sides PL/4, PL/4, and PL/10, respectively, and the certain foreground by eroding them with SEs with sides PL/1, PL/3, and PL/8, respectively. Both algorithms are guided by the extracted upper body mask; therefore, their results are very similar. Their main difference is that Grab Cut can make better guesses in cases of uncertainty and segment large regions loosely defined by the map, whereas GrowCut is more sensitive to the map and more influenced by background seeds. In Fig. 10, for example, both algorithms extract the upper body successfully, but Grow Cut removes the small enclosed regions by the arms, whereas Grab Cut includes them.

Fig. Example of foreground/ background certainty maps and segmentations for (a), (b) Grab Cut and (c), (d) Grow Cut.

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.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

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.

The expected dimensions of the torso are again calculated based on anthropometric constraints, but in a more accurate model. In addition, in order to cope with slight body deformations, we allow the rectangle to be constructed according to a constrained parameter space of highest granularity and dimensionality. Specifically, we allow rotations with respect to rectangle's center by angle φ, translations in x- and y-axes, τx and τ y and scaling in x- and y-axes, sx and sy . The initial dimensions of the rectangle correspond to the expected torso in full frontal and upright view and it is decreased during searching in order to accommodate other poses. The rationale behind the fitting score of each rectangle is measuring how much it covers the UBR, since the torso is the largest semantic region of the upper body, defined by potential upper body coverage (UBC), while at the same time covering less of the background region, defined by potential S (for Solidity). Finally, in many cases, the rectangle needs to be realigned with respect to the face's center (FaceCenter) to recover from misalignments caused by different poses and errors. A helpful criterion is the maximum distance of the rectangle's upper corners (LShoulder, RShoulder) from the face's center (Dsf), which should be constrained. Thus, fitting of the torso rectangle is formulated as a maximization problem.

$$\theta max f(\theta) = \alpha_1 \times UBC(\theta) + \alpha_2 \times S(\theta) + \alpha_2 \times D_{sf}(\theta)$$

$$\theta max f(\theta) = \alpha_1 \times UBC(\theta) + \alpha_2 \times S(\theta) + \alpha_2 \times D_{sf}(\theta)$$

(8)

Where $\theta\theta = (\phi, T_x, T_y, s_x, s_y)$

$\phi, T_x, T_y, s_x, s_y)$

$$UBC(\theta)\theta) = \frac{\Sigma TorsoMask(\theta) \cap UBR}{\Sigma TorsoMask(\theta)}$$

$$\frac{\Sigma TorsoMask(\theta) \cap UBR}{\Sigma TorsoMask(\theta)}$$

$$S(\theta)\theta) = \frac{\Sigma TorsoMask(\theta)}{\Sigma UBR} \frac{\Sigma TorsoMask(\theta)}{\Sigma UBR}$$

$$D_{sf}(\theta)\ D_{sf}(\theta) = e^{\frac{-\left|Max_{D_{sf}} - 1.5 \times PL\right|}{1.5 \times PL}}$$

$$e^{\frac{-\left|Max_{D_{sf}} - 1.5 \times PL\right|}{1.5 \times PL}}$$

Fig. Best torso rectangle with shoulder and beginning of the legs positions.

we estimate the shoulder positions (top corners of the rectangle), and more importantly, the waist positions (lower corners of the rectangle). In turn, waist positions approximately indicate the beginning of the right and left leg legBR = (x,y) and legBL = (x,y), respectively. These points are the middle points of the line segments of the waist points and the point in the center of the line that connects them. Fig. 11 shows a case of a fitted torso and the aforementioned points. Similarly to upper body extraction and the torso rectangle fitting case, we explore hypotheses about the leg positions using rectangles by first creating rectangle masks for the upper leg parts and using them as samples for the

pants color and finally perform appearance matching and evaluate the result. The assumption we make here is that there is uniformity in the color of the upper and lower parts of the pants.

In the case of short pants, where the lower leg parts are naked, the previously calculated skin regions are used to recover them. In order to reduce computational complexity, the size and position of the upper leg rectangles are fixed and adhering to anthropometric constraints and the only free parameter is their angle of rotation with respect to their center φright and φleft . Let Leg Mask(θ) be the binary mask for the two hypothesized leg parts, where θ = (φright,φleft). Every possible upper leg mask is used as a sample of the pants regions, and the leg regions are estimated using the clothes and

Fig. Example legs mask for φright = 0 and φleft = 0

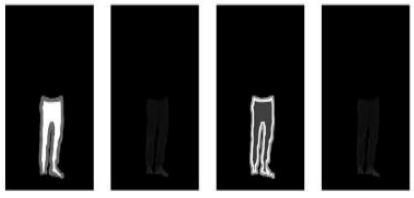


Fig. Example of foreground/background certainty maps and segmentations for (a) and (b) Grab Cut and (c) and (d) Grow Cut.

Skin Detection Process (1)–(5). After the leg potentials are found, the same thresholding process as in the case of the upper body takes place, with the difference that now the expected lower body size is used in (6) (Exp Lower Body Size instead of Exp Upper Body Size), where Exp Lower Body Size = 6 × PL2 . In order to construct the tri map of Grow Cut to perform the refinement process for the leg regions, the leg mask is eroded by a square structuring element (SE) with side PL/4 followed by dilation by a SE with side PL/5 in order to create the uncertainty mask, and for the certain foreground mask it is eroded using a SE with side PL/3. Fig. 13 shows an example. In some cases, thin and ambiguous regions like belts or straps might end up belonging to both the upper and lower body, or in the worst case the background. Most of the time, however, the refinement of the upper and lower regions is able to recover them, and during merging of the two regions they are included in the final outcome.

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 with an appearance model created for each face, to better adapt to the corresponding human's skin color (Fig. 5.2). 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 Grab Cut 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. 5.3.

Each face region FR 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),

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

YCbCr, and Lab colors paces, 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 . The parameters that represent the model are the mean values μij and standard deviations σ ij for each FR and channel j = 1 ... 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:

$$P_{Skin_i}(X) = \prod_{j=1}^{6} \mathcal{N}(X, \mu_{ij}, \sigma_{ij})$$

$$P_{Skin_i}(X) = \prod_{j=1}^{6} \mathcal{N}(X, \mu_{ij}, \sigma_{ij})$$

(1)

Fig. 5.3. Skin detection examples

The adaptive model in general focuses on achieving a high score of true positive cases. However, most of the time it is too "strict" and suppresses the values of many skin and skin-like pixels that deviate from the true values according to the derived probability distribution. At this point, we find that an influence of the skin global detection algorithm is beneficial because it aids in recovering the uncertain areas. Another reason we choose to extend the skin detection process is that relying solely on an appropriate color space to detect skin pixels is often not sufficient for real-world applications. The two proposals are combined through weighted averaging (with a weight of 0.25 for the global model, and 0.75 for the adaptive model). The finest level of image segmentation is used at this point to characterize segments as certain and probable background and foreground. For the certain foreground regions, however, only the pixels with sufficiently high probability in the adaptive model are used as seeds; therefore, as to control their strong influence. In order to characterize a region as probable background or foreground, its mean probability of the combined probability must be above a certain threshold (empirically set to 0.2 and 0.3, respectively). Examples can be seen in Fig. 5.5.

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.

Once the elliptical region of the face is known, we proceed to the foreground (upper body) probability estimation. To better utilize the existing spatial and color relations of the image pixels, we perform multiple level oversegmentation and examine the resulting superpixels. We regard superpixels with color similar to that of the face region as skin and superpixels with color similar to the regions inside torso masks as clothes. With respect to clothes, the size of face's ellipse guides the construction of rectangular masks for the foreground using anthropometric constraints. Our basic assumption is that a good foreground mask should contain regions that appear mostly inside the mask and not outside (background). In other words, we try to identify "islands of saliency," in the aforementioned sense. As opposed to approaches based on pose estimation, we employ simple heuristics to conduct a fast and rough torso pose estimation and guide the segmentation process.

The torso is usually the most visible body part, connected to the face region and in most cases below it. Using anthropometric 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 assumptions 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. This criterion allows for fast inference about the torso's size and location, while relieving the need for the complex task of explicit torso estimation, without sacrificing accuracy.

As discussed, different levels of segmentation give rise to different 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 μi and standard deviation σi of segment Si are estimated for each channel 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

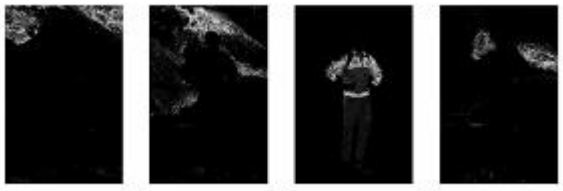


Fig. Example of similarity images for random segments.

Fig. Masks used for torso localization.

In Each Channel Separately:

Example similarity images are shown in Fig. 5. The resulting image that depicts the probability segment Si that is the same color as the rest of the segments is referred to as the similarity image. Similarity images are gathered for all of the different segmentation levels l. Here, we use two segmentation levels in this stage of 100 and 200 super pixels, because they provide a

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

good tradeoff between perceptual grouping and computational complexity

$$P_{SimIm_{li}}(X) = \prod_{j=1}^{3} \mathcal{N}(X, \mu_{ij}, \sigma_{ij})$$

$$P_{SimIm_{li}}(X) = \prod_{j=1}^{3} \mathcal{N}(X, \mu_{ij}, \sigma_{ij})$$

(2)

Sequentially, a searching phase takes place, where a loose torso mask is used for sampling and rating of regions according to their probability of belonging to the torso. Since we assume that sleeves are more similar to the torso colors than the background, this process combined with skin detection actually leads to upper body probability estimation. The mask is used for sufficient sampling instead of torso fitting; therefore, it is estimated as a large square with sides of 2.5PL, with the top most side centered with respect to the face's center. In order to relax the assumptions about the position and pose of the torso, the mask is rotated by 30 ◦ left and right of its initial position (0 ◦) (see in Fig. 5.5). By using a large square mask and allowing this degree of freedom, we manage to sample a large area of potential torso locations. By constraining its size according to anthropometric constraints, we make the foreground/background hypotheses more meaningful.

During the search process, the mask is applied to each similarity image and its corresponding segment is scored. Let Torso Mask be a binary image, where pixels are set to 1 (or "on") inside the square mask and 0 (or "off") outside so that Simile ∩ Torso Mask selects the probabilities of the similarity image that appear inside the mask. Index t = 1, 2, 3 corresponds to a torso mask at angle −30, 0, or 30. Thus, (3) and (4) rate each segment's potential of belonging to the foreground and background, respectively, and (5) combines the two potentials in the form of a ratio as follows:

$$P_{\mathrm{FG}}(S_{li}) = \sum^{|S_{li}|} \mathrm{SimIm}_{li} \cap \mathrm{TorsoMask}_t \quad (3)$$

$$P_{\mathrm{BG}}(S_{li}) = \sum^{|S_{li}|} \mathrm{SimIm}_{li} \cap \overline{\mathrm{TorsoMask}_t} \quad (4)$$

$$\mathrm{TorsoScore}(S_{li}) = \frac{P_{\mathrm{FG}}(S_{li})}{P_{\mathrm{BG}}(S_{li}) + \epsilon}. \quad (5)$$

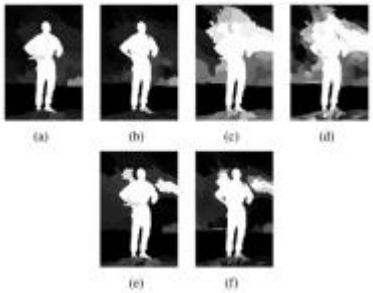


Fig. Segments with potential of belonging to torso. (a), (b) For segmentation level 1 and 2 and torso mask at 0 ◦. (c), (d) For segmentation level 1 and 2 and torso mask at 30 ◦. (e) (f) For segmentation level 1 and 2 and torso mask at −30 ◦.

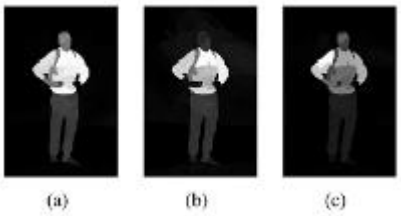


Fig. Aggregation of torso potentials shown in Fig. 7, for torso masks at 0 ◦, 30 ◦, and −30 ◦

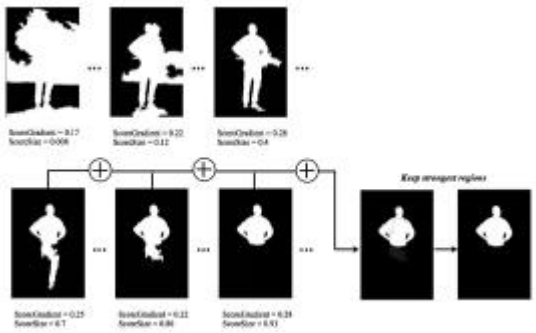


Fig. 5.8. Thresholding of the aggregated potential torso images and final upper body mask. Note that the masks in the top row are discarded.

we can achieve accurate and robust results without imposing computational strain. The obvious step is to threshold the aggregated potential torso images in order to retrieve the upper body mask. In most cases, hands or arms' skin is not sampled enough during the torso searching process,

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

especially in the cases, where arms are outstretched. These segments are superimposed on the aggregated potential torso images and receive the highest potential (1, since the potentials are normalized).

Instead of using a simple or even adaptive thresholding, we use a multiple level thresholding to recover the regions with strong potential according to the method described, but at the same time comply with the following criteria:

1) they form a region size close to the expected torso size (actually bigger in order to allow for the case, where arms are outstretched), and

2) the outer perimeter of this region overlaps with sufficiently high gradients. The distance of the selected region at threshold t (Region) to the expected upper body size (Exp Upper Body Size) is calculated as follows:

$$\text{ScoreSize} = \frac{e^{\frac{-|\text{Region}_t - \text{ExpUpperBodySize}|}{\text{ExpUpperBodySize}}}}{e^{\frac{-|\text{Region}_t - \text{ExpUpperBodySize}|}{\text{ExpUpperBodySize}}}} \quad (6)$$

where Exp Upper Body Size = 11 × PL2 . The score for the second criterion is calculated by averaging the gradient image (Grad Im) responses for the pixels that belong to the perimeter (Region) of Region as

$$\text{ScoreGrad} = \frac{1}{|PRegion_t|}\frac{1}{|PRegion_t|} \quad (7)$$

Thresholding starts with zero and becomes increasingly stricter at small steps (0.02). In each thresholding level, the largest connected component is rated, and the masks with Score Grad > 0.05 and Score Size > 0.6 are accumulated to a refined potential image (see in Fig. 5.8). Incorporation of this a priori knowledge to the thresholding process aids the accentuation of the true upper body regions (UBR). Accumulation of surviving masks starts when Score Size > 0.6 and resulting masks after this point will keep getting closer monotonically to the expected region size. Accumulation ends when Score Size drops below 0.6. The rationale behind this process is to both restrict and define the thresholding range and focus the interest to segments with high potential of forming the upper body segment. The aggregate mask (Aggregate Mask) can now be processed easily and produce more meaningful results. Specifically, we set a final threshold, which allows only regions that have survived more than 20% of the accumulation process in the final mask for the UBR. This process is performed for every initial torso hypothesis; therefore, in the end, there are three corresponding aggregate masks, out of which the one that overlaps the most with the initial torso mask and obtains the highest aggregation score is selected. The aggregation score shows how many times each pixel has appeared in the accumulation process, implicitly implying its potential of belonging to the true upper body segment.

Refinement:

In many cases, the extracted upper body mask is very accurate and can be used as a final result. However, we choose to add an extra refinement step to cope with probable segmentation errors and pixels that manage to survive the multiple thresholding process. One idea that we use here is to give the upper body mask as input to an interactive foreground/background algorithm that requires "seeds" corresponding to the foreground and background. Grow Cut and Grab Cut are used for experiments.

For GrowCut, the uncertain region is constructed by dilating the upper body mask with a SE with sides equal to PL/6, the face's ellipse with a SE with sides equal to PL/10 and the skin regions with a SE with sides equal to PL/12. Possible holes between the face and torso region are also filled. The certain foreground is similarly constructed with erosion instead of dilation, where the sides of the SEs are now PL/4, PL/4, and PL/10, respectively. The rest of

the map is classified as background. For the Grab Cut algorithm, the possible background ground is constructed by dilating the upper body mask, the face's ellipse and skin masks using SEs with sides PL/10, PL/2, and PL/12, the probable foreground is constructed by eroding the masks with SEs with sides PL/4, PL/4, and PL/10, respectively, and the certain foreground by eroding them with SEs with sides PL/1, PL/3, and PL/8, respectively. Both algorithms are guided by the extracted upper body mask; therefore, their results are very similar. Their main difference is that Grab Cut can make better guesses in cases of uncertainty and segment large regions loosely defined by the map, whereas GrowCut is more sensitive to the map and more influenced by background seeds. In Fig. 10, for example, both algorithms extract the upper body successfully, but Grow Cut removes the small enclosed regions by the arms, whereas Grab Cut includes them.

Fig.5.9. Example of foreground/ background certainty maps and segmentations for (a), (b) Grab Cut and (c), (d) Grow Cut.

5.5 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.

The expected dimensions of the torso are again calculated based on anthropometric constraints, but in a more accurate model. In addition, in order to cope with slight body deformations, we allow the rectangle to be constructed according to a constrained parameter space of highest granularity and dimensionality. Specifically, we allow rotations with respect to rectangle's center by angle φ, translations in x- and y-axes, τx and τ y and scaling in x- and y-axes, sx and sy . The initial dimensions of the rectangle correspond to the expected torso in full frontal and upright view and it is decreased during searching in order to accommodate other poses. The rationale behind the fitting score of each rectangle is measuring how much it covers the UBR, since the torso is the largest semantic region of the upper body, defined by potential upper body coverage (UBC), while at the same time covering less of the background region, defined by potential S (for Solidity). Finally, in many cases, the rectangle needs to be realigned with respect to the face's center (FaceCenter) to recover from misalignments caused by different poses and errors. A helpful criterion is the maximum distance of the rectangle's upper corners (LShoulder, RShoulder) from the face's center (Dsf), which should be constrained. Thus, fitting of the torso rectangle is formulated as a maximization problem.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 16
June 2018

$$\theta maxf(\theta) = \alpha_1 \times UBC(\theta) + \alpha_2 \times S(\theta) + \alpha_2 \times D_{sf}(\theta)$$

$$\theta maxf(\theta) = \alpha_1 \times UBC(\theta) + \alpha_2 \times S(\theta) + \alpha_2 \times D_{sf}(\theta)$$

(8)

Where $\theta\theta = (\phi, T_x, T_y, s_x, s_y)$

$\phi, T_x, T_y, s_x, s_y)$

UBC($\theta)\theta) = \frac{\Sigma TorsoMask(\theta) \cap UBR}{\Sigma TorsoMask(\theta)}$

$\frac{\Sigma TorsoMask(\theta) \cap UBR}{\Sigma TorsoMask(\theta)}$

S($\theta)\theta) = \frac{\Sigma TorsoMask(\theta)}{\Sigma UBR} \frac{\Sigma TorsoMask(\theta)}{\Sigma UBR}$

$D_{sf}(\theta)\ D_{sf}(\theta) = e^{\frac{-\left|Max D_{sf} - 1.5 \times PL\right|}{1.5 \times PL}}$

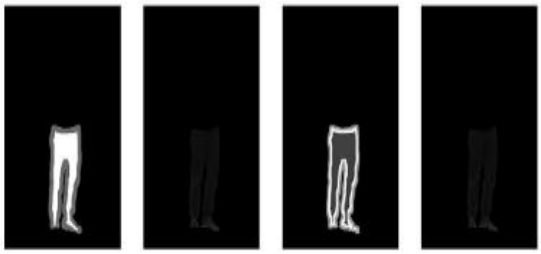
$e^{\frac{-\left|Max D_{sf} - 1.5 \times PL\right|}{1.5 \times PL}}$

Fig.5.10. Best torso rectangle with shoulder and beginning of the legs positions.

we estimate the shoulder positions (top corners of the rectangle), and more importantly, the waist positions (lower corners of the rectangle). In turn, waist positions approximately indicate the beginning of the right and left leg legBR = (x,y) and legBL = (x,y), respectively. These points are the middle points of the line segments of the waist points and the point in the center of the line that connects them. Fig. 11 shows a case of a fitted torso and the aforementioned points. Similarly to upper body extraction and the torso rectangle fitting case, we explore hypotheses about the leg positions using rectangles by first creating rectangle masks for the upper leg parts and using them as samples for the pants color and finally perform appearance matching and evaluate the result. The assumption we make here is that there is uniformity in the color of the upper and lower parts of the pants.

In the case of short pants, where the lower leg parts are naked, the previously calculated skin regions are used to recover them. In order to reduce computational complexity, the size and position of the upper leg rectangles are fixed and adhering to anthropometric constraints and the only free parameter is their angle of rotation with respect to their center φright and φleft . Let Leg Mask(θ) be the binary mask for the two hypothesized leg parts, where θ = (φright,φleft). Every possible upper leg mask is used as a sample of the pants regions, and the leg regions are estimated using the clothes and

Fig.5.11. Example legs mask for φright = 0 and φleft = 0



(a) (b) (c) (d)

Fig.5.12. Example of foreground/background certainty maps and segmentations for (a) and (b) Grab Cut and (c) and (d) Grow Cut.skin detection process (1)–(5). An example mask can be seen in Fig. 12. The hypothesized foreground is the pixels that belong to the leg mask, and background is the rest of the image plus the pixels of the upper body mask, without the pixels below the waist line segment (if any). The leg mask retrieved from each hypothesis is the largest connected component of image segments with color

similar to the hypothesis and the skin regions retrieved in the previous steps. There is no strong need for precise alignment of the masks and the real leg parts, just enough coverage is pursued in order to perform a useful sampling. Thus, the algorithm can recover from slight torso misalignment and performs well in cases of different leg positions, without imposing the computational strain of dense searching using dense mask parameters.

After the leg potentials are found, the same thresholding process as in the case of the upper body takes place, with the difference that now the expected lower body size is used in (6) (Exp Lower Body Size instead of Exp Upper Body Size), where Exp Lower Body Size = 6 × PL2 . In order to construct the tri map of Grow Cut to perform the refinement process for the leg regions, the leg mask is eroded by a square structuring element (SE) with side PL/4 followed by dilation by a SE with side PL/5 in order to create the uncertainty mask, and for the certain foreground mask it is eroded using a SE with side PL/3. Fig. 13 shows an example. In some cases, thin and ambiguous regions like belts or straps might end up belonging to both the upper and lower body, or in the worst case the background. Most of the time, however, the refinement of the upper and lower regions is able to recover them, and during merging of the two regions they are included in the final outcome.

Results

CONCLUSION

We introduced a novel approach for extricating human bodies from single pictures. It is a base up approach that consolidates data from various levels of division so as to find notable locales with high capability of having a place with the human body. The principle part of the framework is the face recognition step, where we appraise the harsh area of the body, develop an unpleasant anthropometric model, and model the skin's shading. Delicate anthropometric requirements manage an effective scan for the most obvious body parts, to be specific the upper and lower body, maintaining a strategic distance from the requirement for solid earlier information, for example, the posture of the body. **ScoreGrad =**

REFERENCES

[1] Solomon, C.J.; Breckon, T.P. Thresholding starts with zero and becomes increasingly stricter at small steps (0.02). In each thresholding level, the largest connected component is rated, and the masks with Score Grad > 0.05 and Score Size > 0.6 are accumulated to a refined potential image (see in Fig. 6.8). Incorporation of this a priori knowledge to the thresholding process aids the accentuation of the true upper body regions (UBR). Accumulation of surviving masks starts when Score Size > 0.6 and resulting masks after this point will keep getting closer monotonically to the expected region size. Accumulation ends when Score Size drops below 0.6. The rationale behind this process is to both restrict and define the thresholding range and focus the interest to segments with high potential of forming the upper body segment. The aggregate mask (Aggregate Mask) can now be processed easily and produce more meaningful (2010). Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab. Wiley-Blackwell. doi: 10.1002/9780470689776. ISBN 0470844736.

[2] Rafael C. Gonzalez; Richard E. Woods; Steven L. Eddins (2004). Digital Image Processing using MATLAB. Pearson Education. ISBN 978-81-7758-898-9.

[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," 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 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. Boykov and M.-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

multiscale superpixels," 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.