ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 14, May 2016

Non interactive gender identification and segmentation of standing humans from static images

T.Dhivya
PG student, Dept of CSE
Mepco Schlenk Engineering College,Sivakasi
dhivya.infotech6@gmail.com

S.Priyadharsini
Asst. Professor, Dept of CSE
Mepco Schlenk Engineering College,Sivakasi
priyadharsini@mepcoeng.ac.in

Abstract— **Segmenting humans from static images is challenging due to various factors such as noise in the image, barricade, inconstant lightning conditions and cluttered background. In general the existing techniques to segment the image need a shape model, templates and user input as seeds. In this paper the method for automatic face detection, gender identification and segmentation of the standing human poses in static images is proposed. The sex classification is done based on facial features in the image. For non-interactive segmentation of the image bottom-up approach is used. The bottom-up approach uses super pixel segmentation to define perceptually uniform region in the image. The skin and face has to be detected which provides a strong indication about the presence of humans in an image, greatly reduces the search space for the upper body. Based on anthropometric measures the shoulder and waist points are estimated which are used to segment the image without any user interaction. The proposed methodology is tested in INRIA person dataset.**

Keywords—anthropometric constraints, SLIC super pixel segmentation, bottom-up approach, Gabor filter

I. INTRODUCTION

In mission vision human body segmentation is required for many applications such as scene understanding, activity recognition, recognizing human actions from static images, determination of the human layout. Many existing approaches have been proposed to perform the specified applications. Interactive methods focus on image segmentation with prior foreground and background seeds which are often given as input from users. These techniques provide better result of segmentation but which require user interactions. However, human body segmentation remains a challenging problem due to the reasons arising from posture variations, inconstant lighting conditions and cluttered backgrounds.

To avoid user interaction in segmenting the human poses, this paper proposes a non-interactive segmentation using bottom-up approach. This approach is one of the ways to segment the image into regions and then identify the image regions that correspond to a single object. Bottom-up approaches use low-level elements, such as pixels or super pixels, and then try to group them into semantic entities of higher levels.

The skin and face is to be detected for identifying a gender and segmenting human from background. Face provides a strong indication about the presence of humans in an image which greatly reduces the search space for the upper body and provides information about facial features. Based on the anthropometric constraints the upper and lower body is estimated to segment the image without user interaction.

II. RELATED WORK

There are many approaches existing for segmenting a human body from the background in still images. The interactive methods referred in [1] expect user input in order to differentiate the foreground and background. Magic Wand an interactive approach starts with a user-specified point or region to compute a region of connected pixels, i.e. all the pixels within some color statistics of the specified region. Intelligent Scissors which allows user to choose an object's boundary with the mouse pointer.

Graph cut [2] provides a convenient language to encode simple local segmentation cues, and a set of robust computational mechanisms to extract global segmentation from the simple local pixel similarity. But it does not run in real time, and it is difficult to expand to other object classes and needs the face detection as the precondition.

Grab Cut [1] an interactive foreground extraction addresses the problem of interactive extraction of a foreground object in a cluttered environment whose background cannot be trivially subtracted. The Grab Cut method does not perform well when a person appears in cluttered backgrounds and need user interaction.

The Picture structure (PS) [4] model is employed based on the original image to exploit more information about the image so that an accurate estimation of articulated poses can be ensured. The computational cost is excessive since the PS model needs to be updated in each iteration.

Top-down approaches are based upon a priori knowledge and use the image content to further refine an initial model. Top-down approaches have been proposed [3] as solutions to the problem of segmenting human bodies from static images. The paper presents an approach to extract human body region without user interaction from color photos, which introduces trimap shape updating into iterated process of Grab Cut image segmentation technique. In this the human body is assumed to be inside a large mask, but due to

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 14, May 2016

the variability of human poses, this assumption often fails, and the sampling may lead to unrecoverable errors.

Baluja and Rowley [18] provided an Adaboost system for gender classification with manually aligned faces. Gutta et al. [17] presented a hybrid classifier for gender determination of human faces that consisted of an ensemble of radial basis functions (RBFs) and decision trees (DTs). that body regions should be comprised by segments that appear strongly inside the hypothesized body regions and weakly in the corresponding background.

The proposed system of automatic face detection, gender identification and segmentation has the modules of upper body, which in turns leads the search for the lower body. Moreover, upper body extraction provides additional information about the position of the hands, the detection of

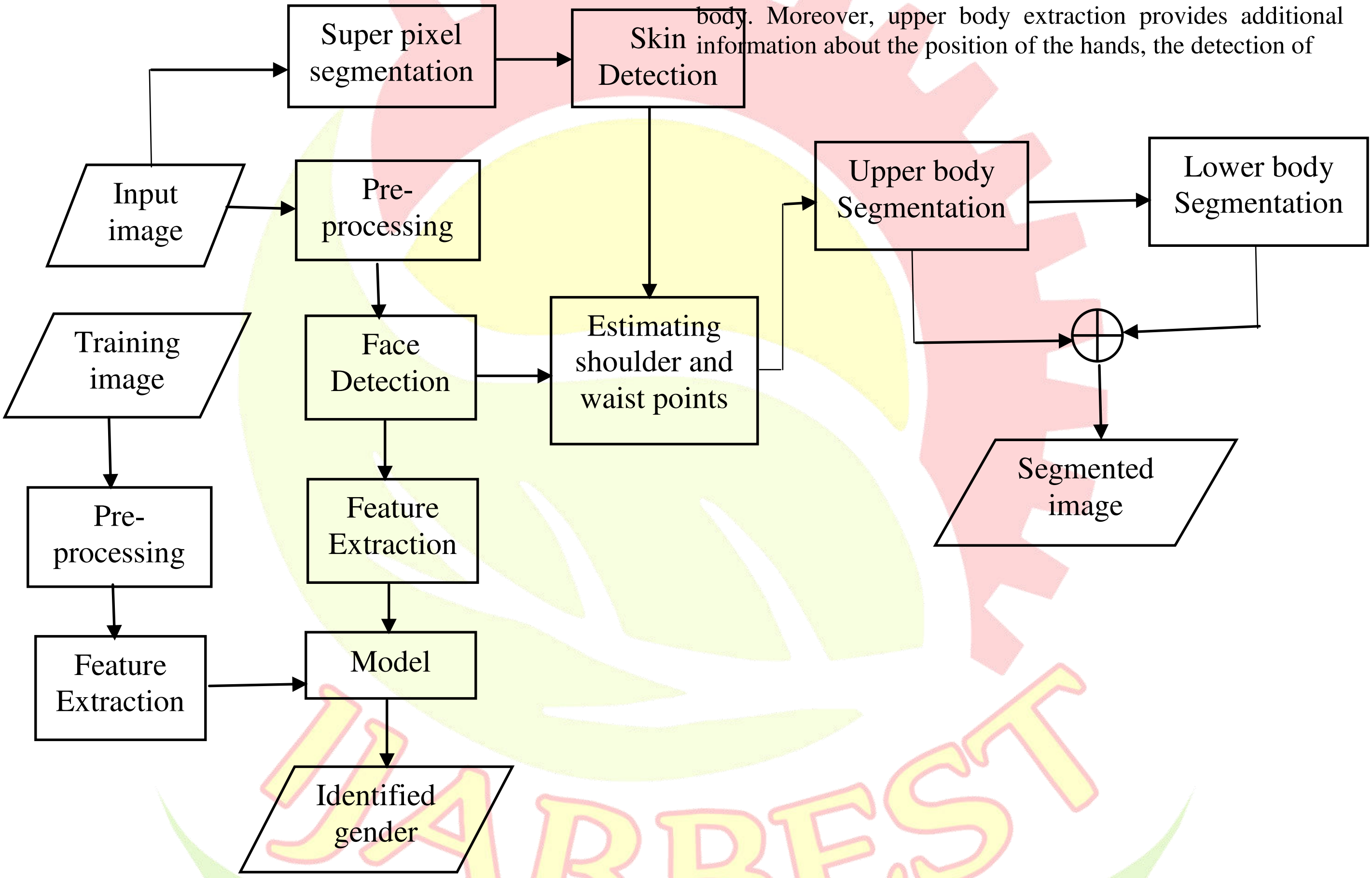

Fig: System Overview

III. PROPOSED SYSTEM

The bottom-up approach, is useful to segment the image into regions and then identify the image regions that belongs to a single object. This segmentation approaches use various image based criteria and search algorithms to find identical segments within the image. The main goal is to detect the object in the image and segment it from the background without the user interaction and to identify the gender for a given input image.

Face dimensions also aid in determining the dimensions of the rest of the body, according to anthropometric constraints. From the detected face regions the Gabor features are calculated and feature vectors are generated. The face information guides the search for the superpixel segmentation, skin detection, face detection, upper and lower body segmentation.

IV. SLIC SUPERPIXEL SEGMENTATION

In general a super pixel is defined as a perceptually uniform region in the image. The merit of using super pixels is computational efficiency. A super pixel identification of an image is helpful to reduce the number of primary images compared to the pixel representation. The SLIC algorithm in [15] estimate the super pixel in the image. The image is first

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 14, May 2016

divided into a grid and the center of each grid tile is then used to initialize a corresponding k-means.

K-means uses the Lloyd algorithm, which alternate the assigned pixels to the closest centers. The only difference compared to standard k-means is that, in k-means each pixel can designate only to the center originated from the neighbor tiles. After k-means has converged, the connected regions are eliminated which are less than minimum region size pixels.

The only parameter of the slic algorithm is k, the desired number of approximately equally-sized superpixel. For color images, the clustering procedure begins with an initialization step k. Initial cluster centers C= [li ai bi xi yi] ^{T} are sampled on a regular grid spaced S pixels apart. To produce roughly equally sized superpixel, the grid interval is S = p, N = k. The centers are moved to seed locations corresponding to the lowest gradient position in a 3 × 3 neighborhood.

Then each pixel, 'i' is associated with the nearest cluster center whose search region overlaps its location. This is the key to speed up the algorithm because limiting the size of the search region significantly reduces the number of distance calculations. This is done using the distance measure D, which determines the nearest cluster center for each pixel.

V. LIGHTNING COMPENSATION

The problem of elimination of non-standard illumination is one of the most complicated problem. The lighting compensation (LC) algorithm is very useful in enhancing and restoring the natural colors into the images which are taken in varying lighting conditions. Therefore, lighting compensation has been used in their skin and face detection algorithms.

The implemented LC algorithm is based on the assumption that the spatial average of surface reflectance in a scene is achromatic.LC algorithm can be defined as follows,

$$S_C = \frac{C_{std}}{C_{avg}} \tag{1}$$

$$C_{std} = \frac{\sum_{i=1}^{i=m}[\max(R_i, G_i, B_i) + \min(R_i, G_i, B_i)]}{2 \times n} \tag{2}$$

$$C_{avg} = \frac{\sum_{i=1}^{i=m}(c_i)_{c_i>0}}{\sum_{i=1}^{i=m}(1)_{c_i>0}} \tag{3}$$

$$n = m\sum_{i=1}^{i=m}(1)_{R_i=G_i=B_i=0} \tag{4}$$

where S_c stands for the scale factor for one specific channel of R, G or B. The C_{std} and C_{avg} are standard mean gray value of the specific channel and the mean value non-black pixels in the same channel.

The m stands for the number of pixels in the image, n stands for the number of non-black pixels in the image. The original image is split into R, G, and B channels for applying the Light compensation algorithm. The output after pre-processing of given image for example is shown in fig 2 (a) and (b).

Fig 2(a): Original Image

Fig 2(b): Pre-processed image

VI. SKIN DETECTION

Skin color information can be considered a very efficient for identifying facial areas under various lighting conditions. Skin color information is also useful in identifying human body and to find the faces in images. As a step towards to segment the human body from the image, the proposed system describe a simple approach to adapt human skin color from the input image to ensure the presence of object.

To ensure the presence of human in the image to segment in a non-interactive manner, the input image has to be converted to lab color spaces. Fig 3: Shows the flow diagram to detect the skin region from the input image. When any image is converted to 'Lab' it will look same as that of an RGB image. An RGB image is easily understood as there are three logical colors. But 'Lab' has a mix of one channel with no color, plus two channels with a dual color combination that have no contrast.

The mean and standard deviation is calculated for image that has been converted from RGB to lab color spaces. The mean and standard deviation is calculated using the following equations,

$$m = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x, y) \tag{5}$$

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 14, May 2016

$$\sigma = \sqrt{\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[f(x,y)-m]^2} \qquad (6)$$

where M and N are rows and columns of the pixels. Using the estimated standard deviation and mean values the threshold of an image has to be evaluated. The threshold is evaluated using the equation,

$$T = a * \sigma_{x,y} + b * m_{x,y} \qquad (7)$$

where T is the threshold and a, b are the non-negative constants which ranges from 0 to 1. Based on the threshold the skin pixels are identified in the image which ensure the object in the image and useful to detect the face of the object.

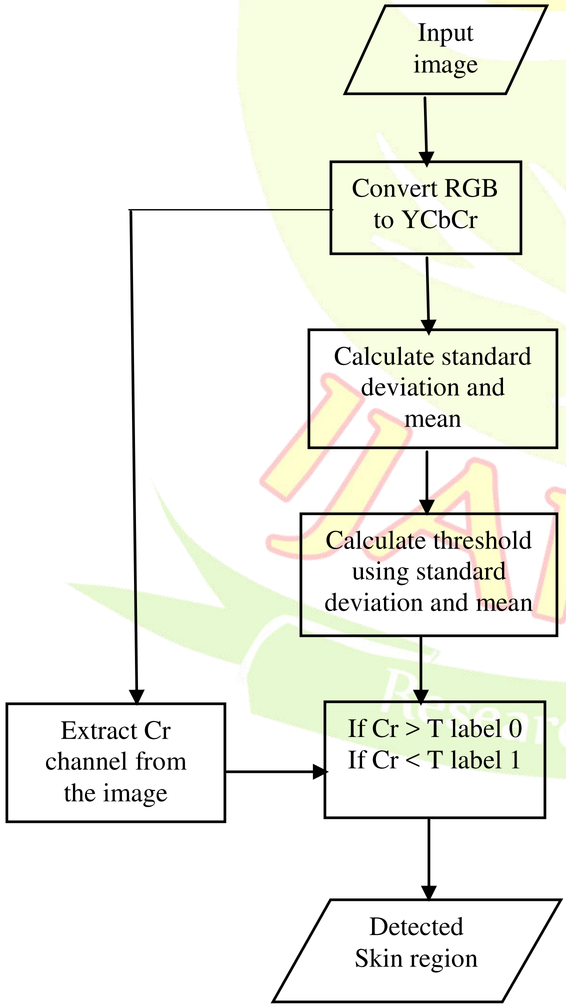

Fig 3: Flow diagram detecting skin region

VII. FACE DETECTION

Face detection is useful to identify the presence of humans in an image, to reduce the search space for the upper body. Face dimensions also helpful in determining the dimensions of the rest of the body, according to anthropometric constraints. Viola–Jones algorithm is used to detect a face in the given single object image.

The first step in the specified technique is computing a set of image features using the integral image. Haar features are used to detect the presence of the computed feature in the given image. The second step in this technique is to choose the features based on AdaBoost. It is a machine learning feature which is useful to identify only the best features among all those estimated features. Once these features are found based on the weighted combination of all these features is used in calculating and deciding any given window has a face or not. The third step in a technique is to perform a cascade of classifiers which helps to reduce computation time.

After fitting a bounding box in the face region, we are able to estimate the locations and sizes of human body parts according to anthropometric constraints. In the case of upper body segmentation, it was the position of the face that aided the estimation of the upper body location.

VIII. FEATURE EXTRACTION AND FEATURE VECTOR GENERATION

The proposed system uses Gabor filters to extract features from the detected face region. The most important advantage of Gabor filters is their invariance to rotation, scale, and translation.

The Gabor filter-based features are directly extracted from the gray-level images. In the spatial domain, a two-dimensional Gabor filter is a Gaussian kernel function modulated by a complex sinusoidal plane wave, defined as:

$$gfilter(x,y) = \frac{(fu)^2}{\pi\gamma\eta}\exp\left[-\left((\alpha)^2\times(x1)^2+(\beta)^2\times(y1)^2\right)\right]\exp[2\pi\times fu\times x1]$$

where,

$$x1 = x\cos\theta + y\sin\theta$$

$$y1 = -x\sin\theta + y\cos\theta$$

$$fu = \frac{f_{max}}{(\sqrt{2})^u}$$

$$\alpha = \frac{fu}{\gamma} \quad ; \quad \beta = \frac{fu}{\eta}$$

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 14, May 2016

$$\theta = \frac{v}{V}\pi \quad , v = 1, 2, V\text{-}1$$

The image is converted into 40 Gabor filters with five scales and eight orientations. The size of the face images used in our experiments is 140*140 pixels. Using forty Gabor filters, the dimension of the feature vector is 140*140*40 = 784,000. Since the adjacent pixels in an image are usually highly correlated, we can reduce this information redundancy by down sampling the feature images resulting from Gabor.

The feature images are down sampled by a factor of four, which means that the feature vector will have a size of 784,000 / (4*4) = 49000. These vectors are then normalized to zero mean and unit variance. The magnitude and real part of the Gabor wavelets in five scales and eight orientation are shown in the fig 4 (a) and (b) respectively.

IX. LOGISTIC REGRESSION

The proposed system uses the classifier based logistic regression. Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function.

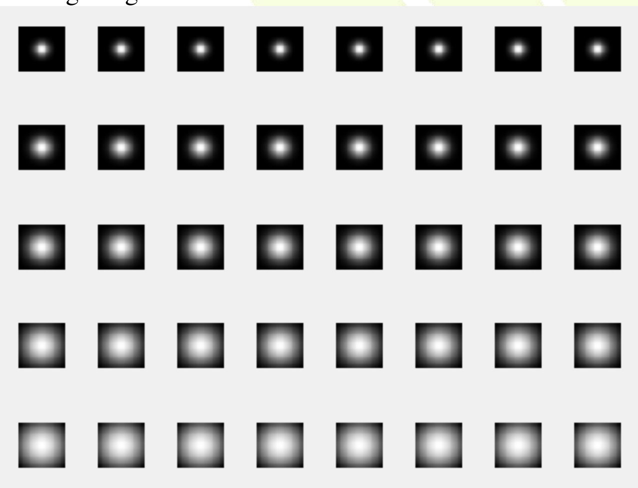

Fig 4(a): An example of Magnitude of Gabor wavelets with 5 scales and 8 orientations

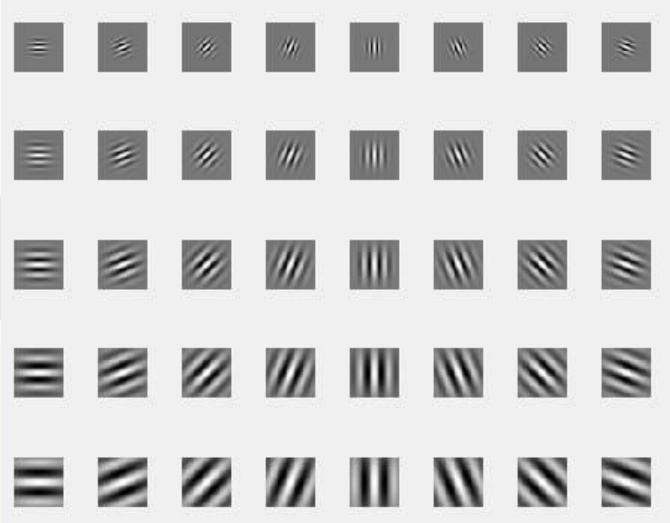

Fig 4(b): An example of real part of Gabor wavelets with 5 scales and 8 orientations

During the probability estimation, the probability of the give face is consider as a male rather than a female based on the generated feature vector X such as,

$$P_{Gender} = \frac{1}{1+\exp(-y)} \tag{9}$$

Where,

$$y = w_0 + w_1 X_1 + \ldots\ldots + w_n X_n$$

$$w_i^1 = \frac{1}{n} \quad , i=1, 2\ldots, n$$

P_{Gender} represents the probability value estimated for a given image and y is the tendency of the face to be male and w represents the weight assigned for each vector. Initially a threshold value of male and female is set as 0.5. When the probability value of the male is greater than the female then the given image is identified as a male. Otherwise when the probability value of female is greater than male then the given image is identified as a female.

X. UPPER BODY DETECTION

In this section, we describe the details for segmenting upper body, which is critical for whole-body segmentation. For a given input image, we first use a face detection method to fit the human face. Once the region of the face is detected for the given image, proceed to the foreground (upper body) estimation. In the case of upper body segmentation, it was the position of the face that aided the estimation of the upper body location. The size of the bounding box in the detected face region can further lead to find the size of body parts according to anthropometric constraints. Anthropometry involves the

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 14, May 2016

standardized measurement of the physical properties of the human body, primarily dimensional descriptors of body size and shape.

The face detection is a crucial step in the proposed algorithm. It is used to define search regions for the other upper body components (shoulder points). In a roughly frontal pose, the shoulder of the person should be positioned directly below the face. From the detected face bounding box the center and the radius has to be evaluated.

Based on the center (xf, yf) and radius (R) of the face the position of the shoulders can be estimated based on the anthropometric constraints measures. The shoulder points are estimated using the equation,

$$sl = (xf - 2R,\ yf + 1.8R) \qquad (10)$$
$$sr = (xf + 2R,\ yf + 1.8R) \qquad (11)$$

where sl and sr are the shoulder left and right points respectively. Once the positions of shoulders are located the waist point has to be identified. Waist points are calculated based on the anthropometric measurements.

The anthropometric measures table [16] states that the minimum measure value from shoulder to waist is 30cm.Using the measures the waist point is estimated.Then the waist point is graphed to the groundtruth image and white pixels above the waist are considered as a upper body mask. The aggregate upper body mask can now be processed easily and produce more meaningful results.

XI. LOWER BODY DETECTION

The algorithm for estimating the lower body part, in order to achieve full body segmentation is similar to the one for upper body extraction. Only change is that the starting points that initiate the leg searching process. In the process of lower body segmentation, it is the upper body that aids the estimation of the lower body's position. The previously estimated upper body is useful for finding the starting point for initiating the leg searching process.

Waist points calculated based on the anthropometric measurements [16] is considered for detecting a lower body mask. The waist point is graphed to the ground truth image and white pixels below the waist are considered as a lower body mask.The detected upper and lower body pixels are compared with the super pixel segmented image and original RGB pixels are identified. Finally the detected upper body and lower body regions have to be merge together to get a final outcome of a segmented human body.

XII. EXPERIMENTAL RESULT

To evaluate the proposed system we used the INRIA person dataset which includes the person's everyday activities. The dataset used to estimate the automatic segmentation of human body from the static image is challenging due to factors like different lightning conditions and various backgrounds. The gender detection can also be separately tested using the FERET database. The FERET database is a standard dataset used for facial recognition system evaluation. The dataset tested includes 2,413 still facial images, representing 856 individuals.

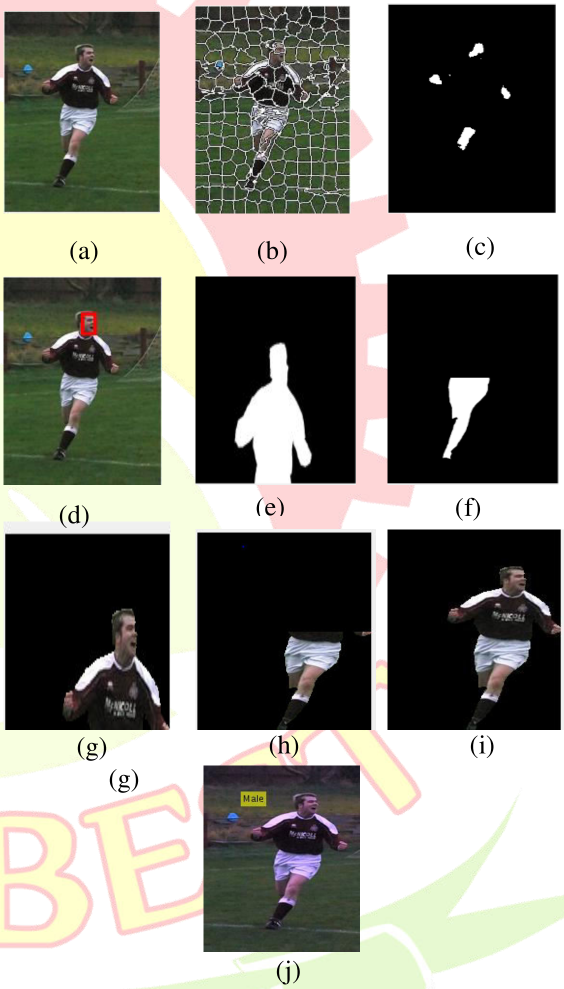

Fig 5: Sample for automatic segmentation of human body from static images (a) Input image (b) Superpixel segmentation (c) Skin detection (d) Detected face (e) Upper body mask (f) Lower body mask (g) Segmented upperbody (h) Segmented lowerbody (i) Segmented human pose (j) Identified gender.

The true positive rate and true negative rate is evaluated for the segmented human poses from the static images. The proposed system focuses on achieving a high score of true positive cases.

The various cases of segmentation of the standing humans from static images is shown in the fig 6.

Input Image Segmented Image

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 14, May 2016

Fig 6: Sample original and segmented images

Using the FERET database the gender identification is verified seperately to Evaluate the accuracy of the gender identification. The sample of images for which the gender identification are shown in the figure 7.

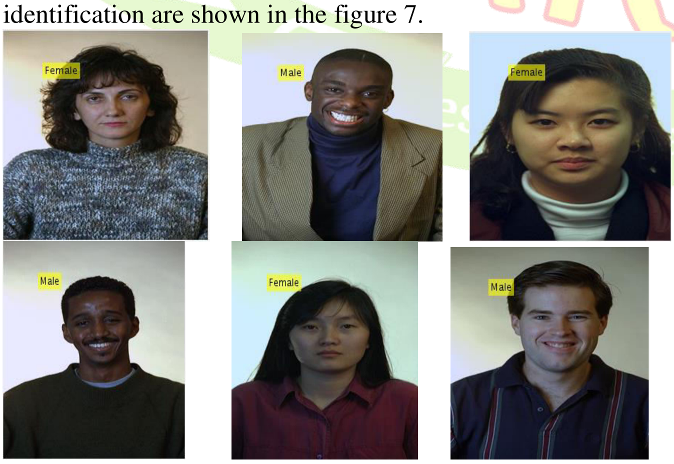

Fig 7: Example of gender identified images

XIII. CONCLUSION

In this paper, we introduced a system for extracting human bodies from static images without any user interaction and identifying gender. It is a bottom up methodology that joins data from various levels of division with a specific end goal to find notable locales with high capability of fitting in with the human body. The main segment of the framework is the face recognition step, where we evaluate the harsh area of the body, and then we develop an anthropometric model, and generate Gabor features to do the sex classification of a given image we have presented logistic regression as classifier in which the classifier is learned by Gabor features. The anthropometric constraints provide an efficient search for the most visible body parts, namely the upper and lower body, avoiding the need for prior knowledge, such as the posture of the human body.

ACKNOWLEDGEMENT

We wish to thank the Management and the Principal of Mepco Schlenk Engineering College, for the support in carrying out this research work. We are extremely thankful and indebted to our beloved Head of the Department Dr. K. Muneeswaran for sharing his expertise, and for his valuable guidance and encouragement extended to us.

REFERENCES

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

[2] Y. Y. Boykov andM.-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.

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

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

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

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

[7] Z. Hu, G. Wang, X. Lin, and H. Yan, “Recovery of upper body poses in static images based on joints

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 14, May 2016

detection," *Pattern Recog. Lett.* vol. 30, no.5, pp. 503–512, 2009.

[8] B. Moghaddam, M.H. Yang, "Learning Gender with Support Faces", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 24, No. 5, pp. 707 – 711, 2002.

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

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

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

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

[13] Athanasios Tsitsoulis and Nikolaos G. Bourbakis, " A Methodology for Extracting Standing Human Bodies From Single Images", IEEE Transactions On Human-machine Systems, Vol. 45, No. 3, June 2015.

[14] H. Kim, D. Kim, et al. "Appearance-based gender classification with gaussian processes," Pattern Recognition Letters 27 (2006) 618–626.

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

[16] R. Drillis and R. Contini, "Body segment parameters,"Tech. Rep. 1166-03, Office of Vocational Rehabilitation, 1966.

[17] Gutta S., Huang J.R.J., Jonathon P. and Wechsler H., "Mixture of experts for classification of gender, ethnic origin, and pose of human faces", in *IEEE Transactions on Neural Networks*, vol. 11, iss. 4, pag.948-960, July 2000.

[18] S. Baluja and H.A. Rowley, "Boosting Sex Identification Performance," Int'l J. Computer Vision, vol. 71, no. 1, pp. 111-119, 2007.

[19] F. Scalzo, G. Bebis, M. Nicolescu, and L. Loss,"Feature fusion hierarchies for gender classification". In Proceedings of International Conference on Pattern Recognition, 2008.