International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Human Body Extraction From Single Images Based on the Multiple Segmentation Algorithm and SPLINE Approach

Manjushri Warad (M.E.)[1], S.S. Mudda[2]

[1]M. S. Bidve Engineering College, Latur, Maharashtra (413512), India

[2]M.Tech.(Communication Systems), Assistant Professor, Bidve Engineering College, Latur, Maharashtra, India

Abstract: ***Human body extraction is a challenging task in the digital image processing and the intense research is carried out in the past years to solve the problems associated with it. The past algorithms partially succeed in achieving the desired results with approximate performance in terms of extracting the exact human body from the associated image on which it lies. In the proposed method a three segment algorithm, namely skin detection, upper body segmentation and lower body segmentation are implemented to extract the exact human body segmentation based on the multiple image segmentation algorithm, the extracted data needs to be perceived in a better way and that is done using the super pixel approach. The experimental results and the contributed SPLINE approach achieved better results than the traditional state-of-art methods.***

Keywords: Human body extraction, Skin detection, Upper body segmentation, Lower body segmentation, SPLINE approach

1. Introduction

The digital image processing deals with the digital content which is perceived by the human visual system (HVS) to understand the digitally captured content in a pleasant manner. The dominance of the digital image processing over the analog image processing has created a way to optimize the various research fields such as medical, robotics, security, sonar, satellite and its implementation paves way to invent the innovative algorithms which in turn helps to develop the applications which helps the user to understand the digitally captured content in a better way than the earlier signal based statistical approach. Although the digital image processing is a primary constituent of the digital signal processing, but without it we cannot perceive the rest of the electro-magnetic constituents. The perception of the remaining digital signal processing contents needs the digital displays and the digital displays is possible only through the digital image processing. This is the reason behind the dominance of the digital image processing over the entire electro-magnetic spectrum.

The complicated background and the different poses of the human are remained as the difficult problems in the human bodies extraction. The process of the human bodies extraction is a different concept from the unmoved objects extraction and it needs the depth research on it to achieve the desired results as expected by using the computer algorithms. The human bodies extraction process is considered as the challenging task in the digital image processing domain. The presence high-level obstacles have the ability to declined the performance measurements to unimaginable level in the human bodies extraction process and the main obstacles which frequently occurs in the extraction of the human bodies are shading occlusions, cluttering artifacts, unrestricted positions and so on.
The human segmentation into various segments is considered as the perfect solution to all existing problems and in this study a bottom-up approach is proposed on static images to perform the desired human body segmentation. The problem in the human body extraction is composed into three segments using the human body segmentation namely face detection, upper body extraction, and lower body extraction. The three segments are interconnected with each other by direct pairwise correlation approach. The face detection is considered as the key aspect in the proposed study as it indicates the presence of the human in the image by detecting the face location based on the skin color which decreases the search space for the upper body and provides the necessary skin color using the skin detection method. The face detection process is carried out based on the face dimensions and the face dimensions are used to determine the dimensions of the rest of the body as per the anthropometric constraints. The above mentioned information us enough to guide the search for the upper body and the upper body search leads to the search of the lower body. The upper body access the additional information of hands and it is important in various applications. The final thing used in the proposed method is t usage of the super pixels which is used to perform the multiple levels of image segmentation.

2. Proposed Methodology Overview

The contributions which are carried out in the proposed study is as follows

- The automatic systems are gaining popularity in various applications due to its various advantages, especially it has ability to reduce the time consumption to a minimum and has the advantage of creating a platform to accomplish the work in an easy way. In the proposed study, automatic segmentation of human bodies is performed on the single images to increase the performance by decrease the runtime complexity.
- The multiple image segmentation is used in the proposed study to get the information and the most important in the proposed study is to combine the collected data from three different segments to achieve the robustness. The pixels are correlated in a visible manner by using the

correlation parameter and the correlation helps to achieve efficiency.

- The proposed method has achieved the best results over the earlier human body extraction algorithms, but still the soft and uncovered body regions are not detected in a reliable manner. The proposed method makes use of the soft anthropometric constraints to cover the uncovered regions and the process is carried out using the permutation.
- The assumptions made about the foreground and background are presented in the intelligent way. The torso region is successfully separated from the sleeve color and the lower as well as upper parts are similar which is further used to extract the total human body from the image. The searching and extracting process is carried out based on both foreground (robust) and background (weekly outside)

3. Problem Definition

- The image is composed of the two regions named foreground and background. The important information in the majority of the images is relied in the foreground while in rare cases the prominent information is present in the background region.
- The human body extraction is not a new concept and the research work on this concept is carried out from decades and the major drawback occurred in the majority of the works is the presence of the background clutter along with extracted human region and in some exceptional algorithms the extraction process is carried out in excellent way but lack of preservation techniques restricts them to use them in real time applications.
- The major problems which include background clutter presence along with the extracted human body, abnormal segmentation, lack of preservation techniques and various other issues are discussed in depth and the proposed study has came up with the innovative solution i.e. the multiple segmentation scheme.
- The three segments based applicative algorithm model is proposed to present a solution to extract the human body in an exact manner, but the condition is the algorithm is applied on the single images and achieves the better result than the existing extraction algorithms. The technology has the ability to improvise itself and the future improvements can introduce the ability to work on the multiple images.

4. Literature Survey

The human face is considered as the interesting aspect in the human body localization and the three factors which is useful for it are the position, dimensions and the color. According to the anthropometric constraints, the lower and the upper body construction models are designed based on the skin color estimation. The extraction of the pose has attained the great interest and the different levels of segmentation granularity are combined to achieve the highest potential. The necessary quantitative and qualitative experiments are performed for human body parts search from the various segments and later they are combined through the joint estimation for attaining better results than the earlier methods [1].

The classification models have great prominence in analysis and in the literature vast amount of research is carried out to classify the significant things with great significance. The majority of the models used for analysis is the classification and hardly 20% are used models are regression models and both the models have their own significance in terms of advantages and disadvantages with respect to applied research field. The emotion plays a vital role in human face recognition and the classifier algorithms are vividly used to find the face and emotion is equipped manner. The KNN classification approach has the ability to detect the human face and match it with the database statistics. KNN classifier collaboration with the Viola Jones face detection (for face detection) and Mel frequency components (for voice verification)has yielded better results than the traditional systems [2].

5. Proposed Method

A. Face Detection

The face detection process is well accomplished by the Viola and Jones method and it maintains the normalized intensity in order to uniform the shape and size of the depicted face region. The face area detection by using the Viola-Jones method is accomplished based on the Haar-like features and the AdaBoost learning algorithm. The Viola and Jones algorithm is an object detection algorithm provides the object detection statistics in the real time.

Color normalization is used to reduce the lighting effect because the normalization process is actually a brightness elimination process. Input image of $N_1 \times N_2$ pixels represented in the RGB color space, $X = \{X^{n_3}[n_1, n_2] \,|\, 1 \leq n_1 \leq N_1, 1 \leq n_2 \leq N_2, 1 \leq n_3 \leq 3\}$, the normalized values, $X_{norm}^{n_3}[n_1, n_2]$, are defined by

$$X_{norm}^{n_3}[n_1, n_2] = \frac{X^{n_3}[n_1, n_2]}{\sum_{n_3=1}^{3} X^{n_3}[n_1, n_2]} \quad (1)$$

Where $X_{norm}^{n_3}[n_1, n_2]$ for $n_3 = 1,2,3$ corresponding to red, green, and blue (or R, G, and B) components of the image X.

B. Multilevel Image Segmentation

Going through a single pixel for image processing is a very complex method of image processing in real time scenarios. Image compression or any image processing operations produce a severe effect at the pixel level. Mostly filters and some algorithms are used to collect (extract) the data information from the pixel. Compared to single image a group of image gives more information about abstract structure and shape.

In this implementation, we are operating on group of pixels by using image segmentation operation which provides more information.

There are different image segmentation operations, but we have to select only one depending on following criteria

1) We have to preserve strong edges in the image
2) We have to produce segments with relatively uniform size.

Uniformity in size is essential since it limits the calculation from being deceived by over segmenting nearby picture patches of high entropy

Image segmentation method is the basic approach to process the pixels in meaningful groups. However, numerous image segmentation approaches have been proposed in digital image processing, the effective technique is to be selected based on the following criteria. Firstly, the image segmentation at the boundaries must be strong to differentiate the regions as well as to preserve the edges. Secondly, another desirable attribute is to produce segments of uniform size. The above approaches faces problems in such a way that utilize low-level image cues and, thus, their results cannot guarantee compliance with the various and subjective human interpretations. Loss of content at the edges as well as its little bit difficult to segment the image in a uniform order to analyze a particular region.

Mori et al. have proposed a method called entropy rate superpixel segmentation (ERSS) algorithm to use superpixels and normalized cut for segmentation of an image instead of working on pixels. This method provides a tradeoff between the accuracy as well as computational complexity. The proposed approach is based on two components such as entropy rate of a random walk on a graph and a balancing term. In general, superpixels are based on multilevel segmentation like 100 to 500 superpixels depending upon the flexibility. For skin detection, the image is segmented to 500 superpixels in order to differentiate the skin region and skin like regions and to extract only the skin region from an image.

C. Skin detection

The problem to detect the skin regions like face, torso, hands in images and videos is skin tone variations due to illumination and ethnicity. Even limbs do not contain enough information to recognize them easily. The proposed global detection technique is combined with an appearance model for better results of human skin color. The appearance model will provide the differences between the skin and skin like pixels. Grab cut algorithm is used to differentiate the uncertain and certain regions and segmentation cues are used to create regions of uncertainty. Anthropometric constraints and body connectivity will eliminate the false positives.

The adaptive model highly focuses on achieving a high score of true positive cases. It is too "strict" and will differentiate the values of many skin and skin-like pixels and suppresses the values. In this model global detection algorithm is useful because it will recover the uncertain regions and also detect the skin regions. Detecting skin pixels and color space is not sufficient for real-world applications. So the adaptive model algorithm is combined with a global detection algorithm through weighted averaging. (The weights are 0.25 for global detection algorithm and 0.75 for adaptive model). In the final level image segmentation is done to obtain the foreground and background regions of an image. In the adaptive model the pixels with high probabilities are taken from the certain foreground regions and these pixels only used as seeds. In order to make probable foreground and background the combined probability should have a certain threshold level (i.e. 0.2 to 0.3respectively).

D. Upper Body Segmentation

The most important and initial step of our methodology is detection of the facial region, then it will process the rest of the body parts. Here three steps are followed to detect the face region information. The first step is the person's face color should match with the rest of the visible skin areas and make the person adapted to skin detection process. The second step is location of the face region, it provides strong cue about the rough location of the torso. Third one is the size of the face region will provide the guidelines for the size of the body parts with reference to anthropometric constraints. Previously detection of the face region can be done by using Viola-Jones face detection algorithm for both front and side views. Face detection is a keystone of our method. So we filter the results of previously mentioned method using the face detection algorithm.

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

$$P_{simIm_{li}}(X) = \prod_{j=1}^{3} \mathcal{N}\left(X, \mu_{ij}, \sigma_{ij}\right) \qquad (2)$$

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.

Our approach has the advantages of taking different perceptual groupings into account and being able to alleviate the need for accurate torso mask estimation, by conjunctively measuring the foreground and background potentials. The fact that we use super pixels in the computations makes comparisons more meaningful, preserves strong boundaries, and improves algorithmic efficiency. Results may be improved by adding more segmentation levels and masks at different sizes and locations, but at the cost of computational complexity.

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 thresholdt (Region t) to the expected upper body size (ExpUpperBodySize) is calculated as follows:

$$ScoreSize = \frac{-|Region_t\_ExpUpperBodySize\,|}{ExpUpperbody} \qquad (3)$$

where ExpUpperBodySize=11×PL 2 . The score for the second criterion is calculated by averaging the gradient image (GradIm) responses for the pixels that belong to the perimeter (PRegiont)of Regiont as

$$ScoreGrad = \frac{1}{|PRegion_t|} \sum^{|PRegion_t|} GradIm \cap PRegion_t \qquad (4)$$

E. 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 angleφ, translations in x- andy-axes,τx andτy and scaling inx- andy-axes,sx andsy. 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 potentialS(for Solidity). Finally, in many cases, the rectangle needs to be realigned with respect to the face's center (Face Center) 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 constrained. Thus, fitting of the torso rectangle is formulated as a maximization problem

$$\theta max f(\theta) = \alpha_1 \times UBC(\theta) + \alpha_2 \times s(\theta) + \alpha_3 \times D_{sf}(\theta) (5)$$

where TorsoMask(θ) is the binary image, where pixels inside the rectangle rTorsoMask(θ) are 1, else 0; UBR is the binary image, where pixels inside the UBR are 1, else 0; a1,a2,a3 are weights, set to 0.4, 0.5, and 0.1, respectively

6. Spline Detection

In the mathematical field of numerical analysis, spline detection is a form of interpolation where the detector is a special detection is often preferred over polynomial detection because the detection error can be made small even when using low degree polynomials for the spline. Spline interpolation avoids the problem of Runge's phenomenon, in which oscillation can occur between points when interpolating using high degree polynomials. Splines have multiresolution properties that make them very suitable for constructing wavelet bases and for performing multi-scale processing

Splines are piecewise polynomials with pieces that are smoothly connected together. The joining points of the polynomials are calledknuts. For a spline ofdegree n, each segment is a polynomial of degree n, which would suggest that we need nf 1 coefficients to describe each piece. However, there is an additional smoothness constraint that imposes the continuity of the spline and its derivatives up to order (n-1) at the knots, so that, effectively, there is only one degree of freedom per segment

7. Discussion

1) It can automatically separate the human body from the original image.
2) It gives a more accurate image from the original image and it can also give more accurate images even from a complex scenarios.
3) In this method we take cues from the levels of segmentation.
4) By taking the foreground and background of an image, we can reduce the need of an exact mask fitting and dense searching.
5) We allow the masks to be large according to measurements of a human body limitation, by considering the background and foreground of an image. So that we may perform perfect sampling in fewer steps, we demonstrate how soft anthropometric constraints can guide and automate the process in many levels.
6) The main complexities are the segmentation algorithm.

8. Results and Analysis

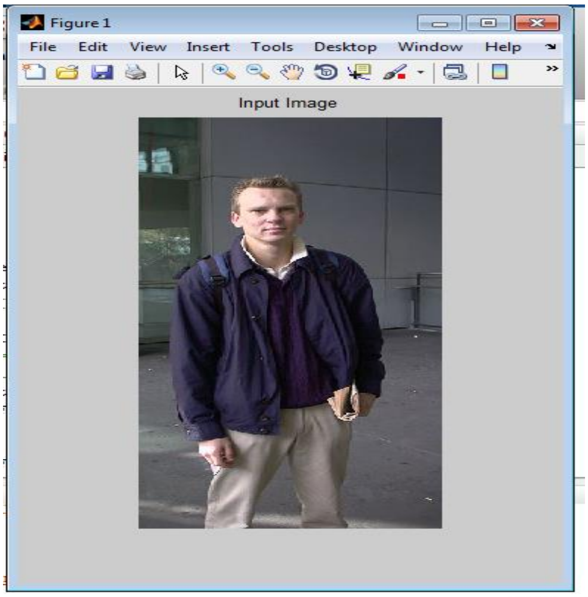

Figure 1: Input Image

Analysis: The input image has taken as shown in above figure which is processed for human body segmentation using the proposed algorithm in further processing

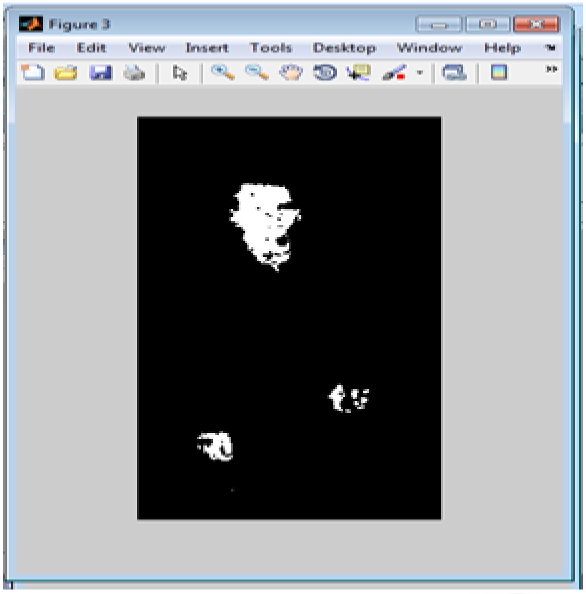

Figure 2: The proposed method result

Analysis: The skin detection result is shown in above figure which is obtained by using the Viola Jones method

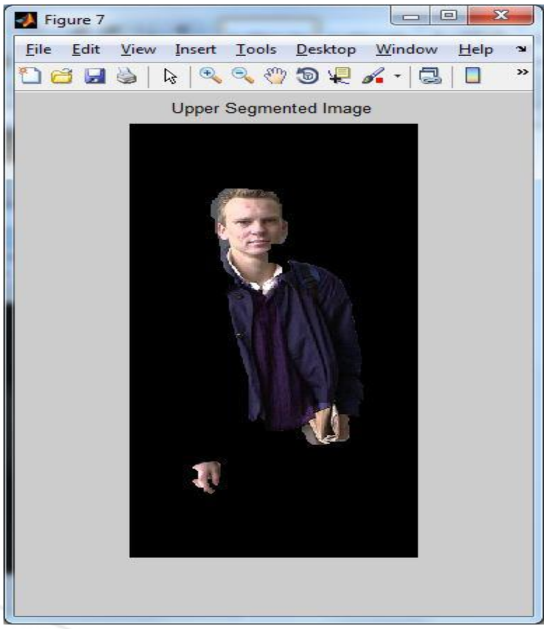

Figure 3: Upper segmented image

Analysis: The upper segmented image is performed using the different parameters. The torso region is separated from the facial region. The upper body segmented image is identified the hand-sleeves apart from the skin detection.

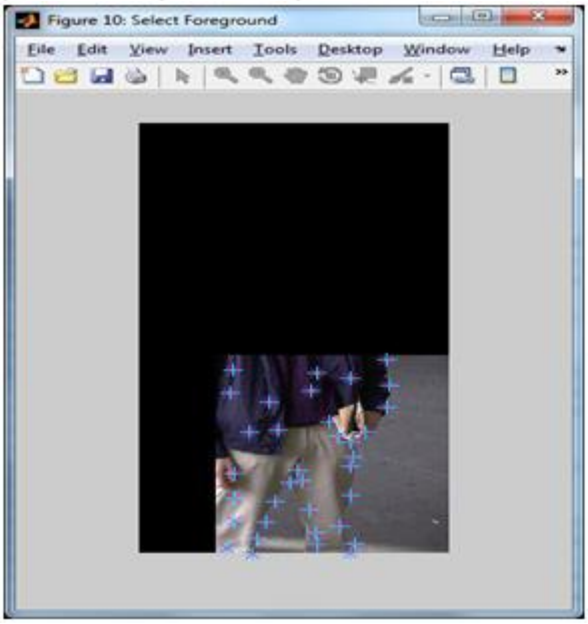

Figure 4 (a): The Foreground Selection

Anslysis: The foreground selection is consider as the important operation in the proposed study and its implementation can be done in above shown format.

International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

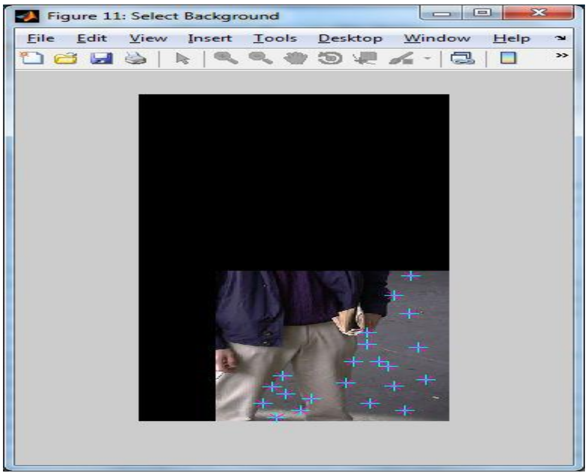

Figure 4 (b): The Background Selection

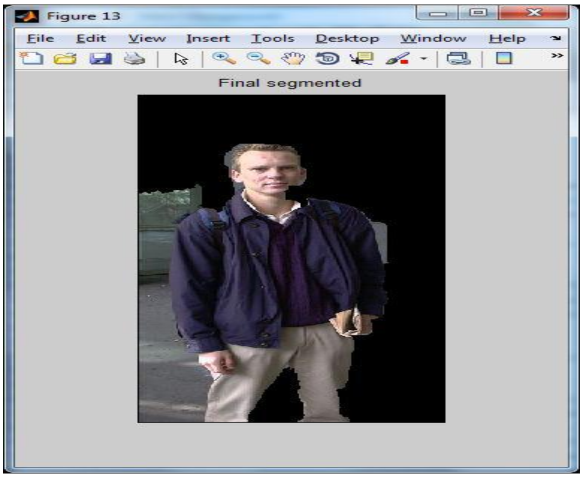

Figure 5: Final segmented (Proposed)

Anslysis: The Background selection is consider as the important operation in the proposed study and its implementation can be done in above shown format.

Analysis: The final segmented human body image is an outcome of the multiple image segmentation algorithm. The proposed shows better results than the earlier methods.

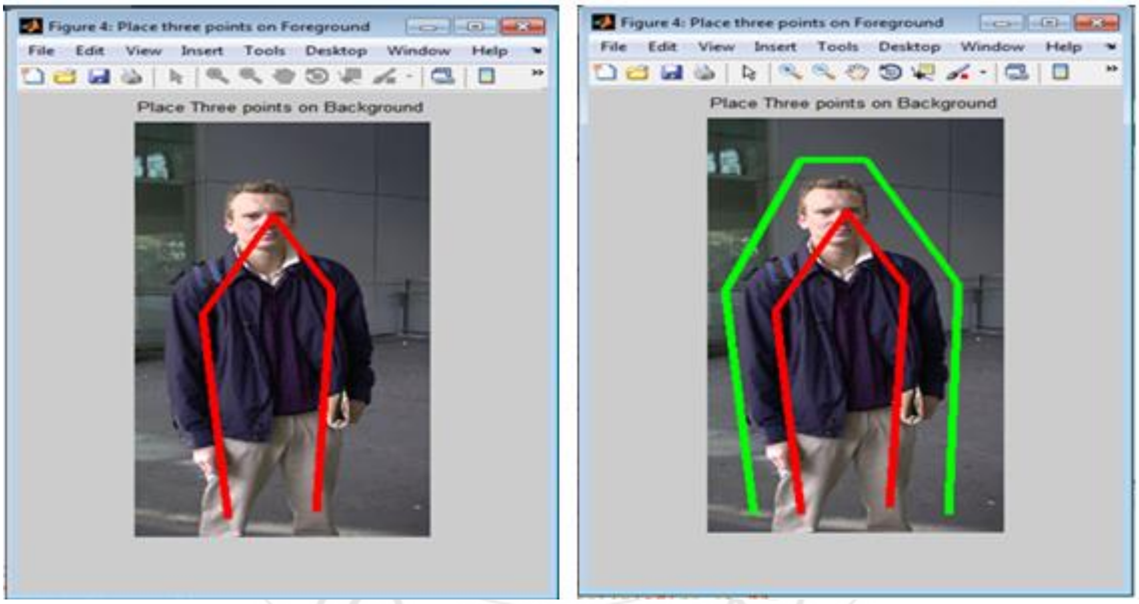

Figure 6: Final segmented (Proposed)

Analysis: The placing of the three points on the background can be perceived in the above two figures and the color representations of red shows the initial points pacing while the red & green shows the final points placing

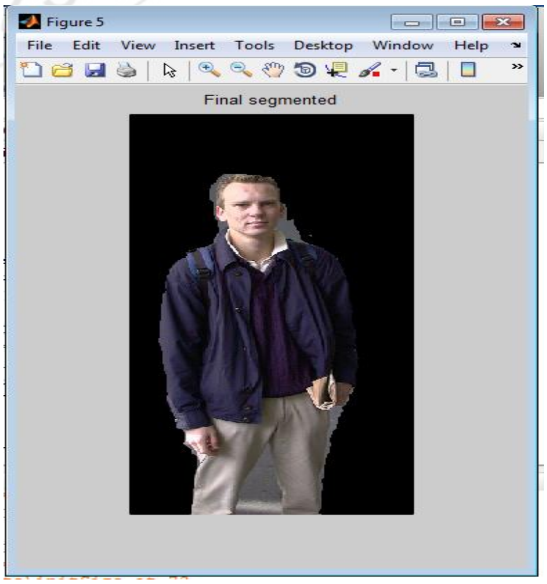

Figure 7: Final segmented (Extension)

Analysis: The final segmented human body image is an outcome of the SPLINE algorithm. The proposed shows better results than the earlier methods.

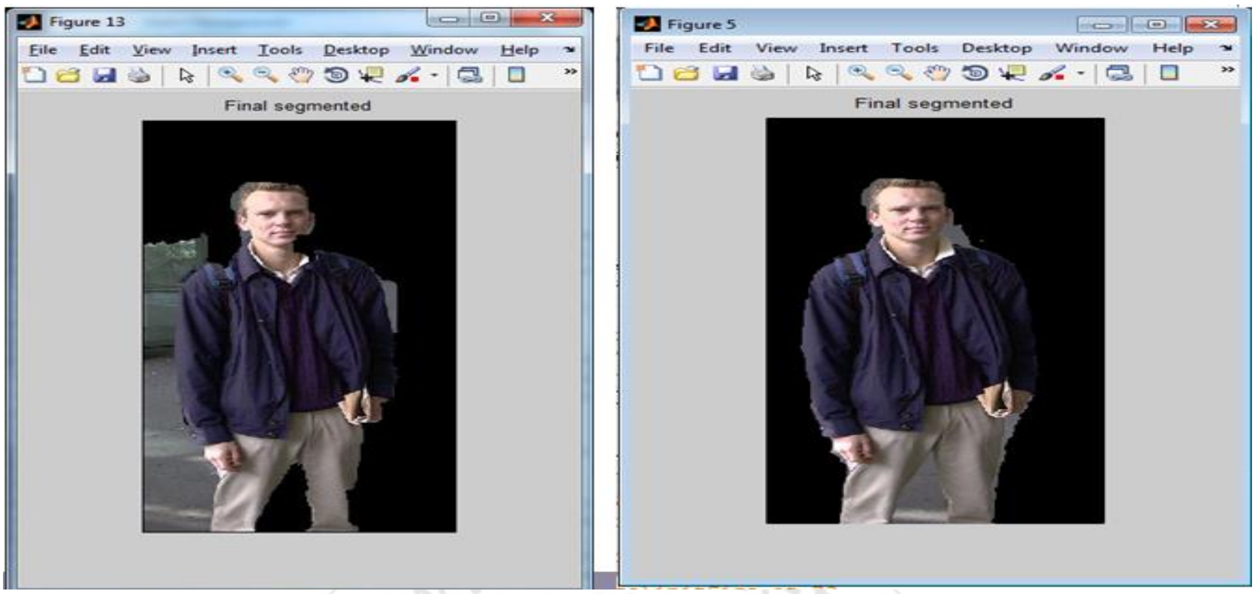

Figure 8: Final segmented (Proposed & SPLINE extension)

9. Conclusion

The human body extraction from the single images is performed in this study using the multiple image segmentation algorithm. The various algorithms such as Viola-Jones method for face detection, Skin detection algorithm, Upper body segmentation, Torso masking and finally lower body segmentation are are implemented in this paper to achieve the better results than the traditional methods. The SPLINE algorithm is contribution work carried out with the proposed method to get much more better results.

References

[1] Athanasios Tsitsoulis, Member, IEEE, and Nikolaos G. Bourbakis, Fellow IEEE", A Methodology for Extracting Standing Human Bodies From Single Images" Vol 45,No 3,JUNE 2013 IEEE.
[2] Dolly Reney ,Dr.NeetaTripaathi Electronics " An Efficient Method to Face and Emotion Detection ",2015 Fifth International Conference on Communication Systems and Network Technologies.2015 IEEE.
[3] Guan-Chun Luh, "Face Detection Using Combination of Skin Color Pixel Detection and Viola-Jones Face Detector ", 16 July 2014, IEEE.
[4] Guan-Chun Luh, "Face Detection Using Combination of Skin Color Pixel Detection and Viola-Jones Face Detector ", 16 July 2014, IEEE.
[5] Xiaohua Wang, Xi Zhang, Jingliang Yao Hangzhou Dianzi University Hangzhou, China, "Skin Color Detection under Complex Background", August 19-22-2011 IEEE.
[6] JatinChatrath, Pankaj Gupta, PuneetAhuja , Aryan Goel, ShaifaliM.Arora" Real Time Human Face Detection and Tracking" ,2014 IEEE.
[7] Min C. Shin Univ of North Carolina at Charlotte, Kyong I. Chang Univ of Notre Dame Leonid V. Tsap Lawrence Livermore National Lab," "Does Colorspace Transformation Make Any Difference on Skin Detection?",2002 IEEE.
[8] Weilong Yang, Yang Wang, and Greg Mori ," Recognizing Human Actions from Still Images with Latent Poses", Canada, 2010 IEEE.
[9] Athanasios Tsitsoulis and Nikolaos Bourbakis ,"A Methodology For Detecting Faces From Different Views", ATR Center, Engineering College, Wright State University, Dayton, Ohio,2012 IEEE.
[10] Yihang Bo, Beijing Jiaotong, China Charless C. Fowlke, Irvine "Shape-based Pedestrian Parsing",2013 IEEE.
[11] MykhayloAndriluka, Stefan Roth, and BerntSchiele,"Pictorial Structures Revisited: People Detection and Articulated Pose Estimation", Department of Computer Science, TU Darmstadt,2009 IEEE.