An Effective Approach for Detecting Human Torso from Static Images

[1]Merin Mariam Mathew [2]Sajin Salim
[1]P.G Scholar [2]Assistant Professor
[1] Electronics and Communication Engineering Department
[1, 2]TKM Institute of Technology,Karuvelil,Kollam

Abstract---Torso detection in static images is a challenging and important problem. Locating the torso region of a human body is difficult due to its cluttering background and various poses. Previous works on torso detection usually demand a time-consuming training phase for complex shape-matching processes. In this paper, a new methodology for detecting the torso portion of the human body is introduced. In this approach, a face detection algorithm is used to locate the face region as a priori and then the torso region is extracted by effectively combined the input image segments and the boundary detection algorithm along with the detected face.

Keywords: Normalized cuts segmentation, Pose estimation, torso detection**.**

I. INTRODUCTION

Torso detection plays an important role in human pose estimation, since torso connects most other body parts (arms, legs and head) together. However, locating a torso region is difficult due to cluttering background and various poses.

A number of schemes have been proposed in past few years for solving this problem. In the exemplar-based approach an exemplar pool should be constructed first. The exemplar pool includes a lot of contour models upon which the joint locations were marked previously are stored. Then, the joint locations of the exemplars can be transferred to test images by shape matching.However, it is impossible to construct an exemplar pool covering all pose variations for shape matching. Gang *et al*.[2] utilize line segments to assemble torso shape, and Ren*et al*.[3] extract the torso region based on a pair of parallel line segments. However, these edge-based methods may not work well when lots of line segments are observed near the torso due to complex backgrounds. The work of Mori*et al*.[4]first searched all combinations of normalized cuts segments that satisfy a scale constraint and then classified these candidates with a set of cues. However their torso result may be broken or contain background. Hu *et al.*[5] detected torso on dominant colors generated by using the k-means clustering algorithm. However the dominant colors are not reliable in appearance variation and cluttered backgrounds.

In this paper, a new methodology of torso detection is introduced. Firstly the head position is obtained by face detection and the input image segments are obtained by normalized cuts algorithm. Then the torso region is extracted by effectively combined these along with the detected boundary. In this scheme, the normalized cuts segments are grouped into torso candidate region based on the bounding box along different orientations. In the combining procedure, three cues are employed to select the best candidate as torso: area probability, location and contour probability. Based on the three cues torso can be estimated.

The remainder of this paper is organized as follows. Section II describes the new methodology of the new torso detection system. Section III describes the face detection technique used in this work. Section IV explains the normalized cuts segmentation algorithm and Section V about the algorithm for boundary detection. Section VI presents the experimental results and discussion while conclusion is stated in Section VII.

II. METHODOLOGY

Torso detection suffers from variations in backgrounds and poses. The result of face detection is used to locate the face region of the input image. The torso is close to the head, and most of the time it is under the head. Then the input image is segmented by using a normalized cuts segmentation algorithm and then the boundary is detected to constrain the segment grouping in to torso region. Then the torso portion of the human body is obtained by combined the face detection algorithm along with the normalized cuts segments and the detected boundary.

Figure 1 shows the methodology of torso detection in this work

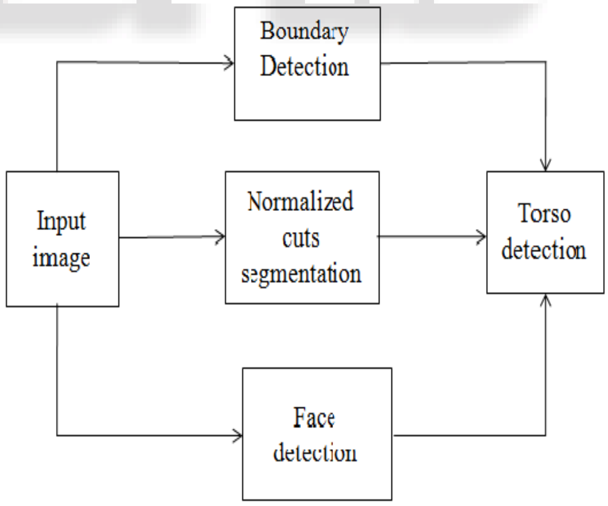


Fig. 1: Methodology for torso detection

III. FACE DETECTION

Face detection [6] is a computer technology that determines the locations and sizes of human faces in digital images. It detects facial features and ignores anything else. Face detection can be regarded as a specific case of object-class detection. In object-class detection, the task is to find the locations and sizes of all objects in an image that belong to a given class. Here the face detection is done by using a visual object detection algorithm called Viola Jones algorithm, which is capable of processing images extremely rapid and achieving high detection rates.

There are three main contributions in this object detection framework. The first is the introduction of a new image representation called the "Integral Image" which allows the features used by our detector to be computed very quickly. The second is a learning algorithm, based on AdaBoost, which selects a small number of critical visual features from a larger set and yields extremely efficient classifiers. The third contribution is a method for combining increasingly more complex classifiers in a cascade manner, which achieves increased detection performance while radically reducing the computation time. This cascade structure allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions.

In the detection phase of viola-Jones algorithm, a window of the target size is moved over the input image, and for each subsection of the image the rectangular feature is calculated. This difference is then compared to a learned threshold that separates non-objects from object.

IV. BOUNDARY DETECTION

The boundary of torso indicates strong global information that can constrain the segment grouping into torso region. Here, the boundary is detected by using a morphological edge detector. The input image is converted into gray level image for the boundary detection. Then a morphological operation called erosion is performed on input image. It shrinks or thins object in the input image by removing pixels on object boundaries. Here erosion is performed by using a structuring element, which is a rectangular array of pixels containing values 1or 0.The value of the output pixel after erosion is the minimum value of all pixels in the input pixels neighborhood. After performing erosion the next step is to extract the boundary of the input image. The boundary region is extracted from the input image by subtracting the eroded image from the input image

V. NORMALIZED CUTS SEGMENTATION

Normalized cuts algorithm [7] is used for segmenting the input image, for detecting the torso portion of the human body. In the normalized cuts algorithm, first a weighted graph is constructed from the input image by taking each pixel as a node and each pair of pixels by an edge. The weight on the edge should reflect the likelihood that the two pixels belong to one object. Using the brightness value of the pixels and their spatial location, the weighted graph is constructed. The weight matrix can be calculated as:

$$w_{ij} = e^{\frac{\left|F_{(i)}-F_{(j)}\right|^2}{\sigma_1^2}} * e^{\frac{\left|X_{(i)}-X_{(j)}\right|^2}{\sigma_2^2}} \qquad (1)$$

Where $F_{(i)}$ and $F_{(j)}$ are the brightness value of nodes i and j and $X_{(i)}$ and $X_{(j)}$ are their corresponding spatial locations. The terms σ_1 and σ_2 are varying parameters in the weight matrix calculation. The value of σ_1 is chosen smaller compared to σ_2.

In the next step, the eigen vectors are computed with the smallest eigen values of the system. After the eigen vectors and eigen values are calculated then the eigen vector with second smallest eigen value is used to bipartition the graph. In the ideal case eigen vector should only take on two discrete values and the graph partitioning depends on the sign of the values but here the eigen vectors took continuous values and chose a splitting point to partition the graph. Here the median value of the second smallest eigen vector was chosen as the splitting point. Hence the graph is broken into two pieces. After the graph is partitioned then again performed the algorithm and repartitioned the graph.

After the normalized cuts segmentation, image segments are obtained and then group these normalized cuts segments into a torso region based on the bounding box, where the bounding box is generated according to the response of face detection. After the face is detected the center of the head is identified. Then based on the center of the head region, according to the height and width of the head an appropriate torso region is recognized. In the combining procedure, three cues are employed to select the coarse torso: area probability, location probability, contour probability

A. Area probability:

Suppose there are N bounding boxes relative to the face and L_i segments overlapped with i[th] bounding box region as R_i the area of the j[th] segment overlapped with R_i as $S_{i,j}$, and the overlapped areas as $o_{i,j}$, j=1,2.....$L_{i..}$ Then area probability $AP_{i,}$ which indicates the jth segment under i[th] bounding box belonging to torso or not, is defined as

$$AP_{i,j} = \left(\frac{O_{i,j}}{S_{i,j}}\right)^{\alpha}\left(\frac{O_{i,j}}{R_i}\right)^{\beta} \qquad (2)$$

Where parameters α and β are the weighting coefficients to control the importance of these two terms.

B. Location probability:

Location probability describes the likelihood of each pixel in a segment unit belonging to the given bounding box candidate. Therefore, the contribution of the segment unit to the candidate can be estimated by the location cue. Let *w* and *H* be the width and height of the bounding box regions, respectively. Given a pixel (x,y), we define

$$\rho_i(x,y) = \left[\left(\frac{d_{i(x,y)}\sin\theta_i}{W/2}\right)^2 + \left(\frac{d_{i(x,y)}\cos\theta_i}{H/2}\right)^2\right]^{1/2} \qquad (3)$$

Where θ_i is the vertical angle of the i[th] bounding box region, and $d_i(x,y)$ is the Euclidean distance from pixel (x,y) to the centre of the bounding box region. Then Location probability LP of $S_{i,j}$ is defined as:

$$LP_{i,j} = \sum_{(x,y)\in S_{i,j}} exp\left(-\left|\frac{p_{i(x,y)}}{W}\right|^{\gamma}\right) \qquad (4)$$

Where parameter γ is the weighting factor.

C. Contour probability:

Contour probability of a region is defined as the probability along its boundary. The boundary of torso constrained the segment grouping into torso region. Based on the above mentioned cues torso region is estimated. Given a bounding box region, all segments that are overlapped with the bounding box region is found out without considering the head region. For each such segment, the area and location probability is computed. Once a segment is added to the torso region, the contour probability is computed and recomputed to constrain the unlimited increase in torso.

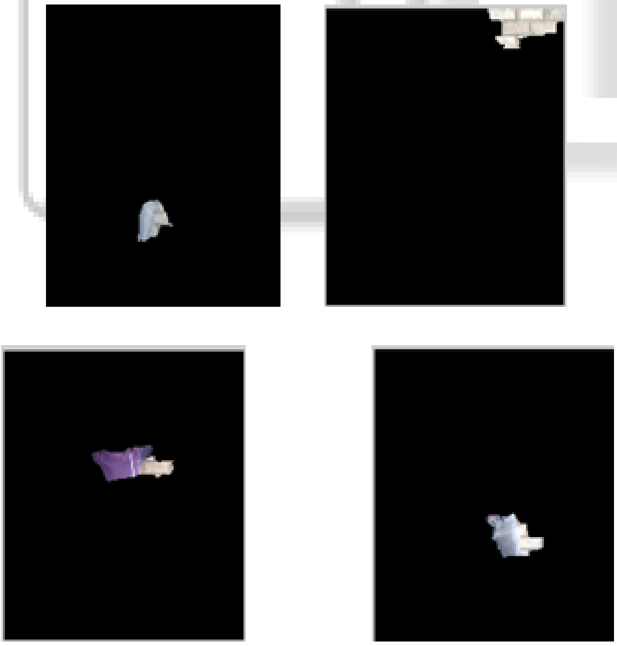
The segment which has maximum contour probability in a bounding box is considered. Then the region inside that bounding box among all other boxes is considered as the torso region.

VI. RESULTS

To evaluate the new methodology, the images of different background, poses and illumination are taken. Simulation is done in MATLAB R2013a. Figure 2 shows the simulated results. Input image to which torso is detected is shown in Figure 2 (a). Figure 2(b) shows the face detected output and the normalized cuts segments are shown in 2 (c) and the Figure 2(d) shows the boundary detected output. Detected torso portion of the human body is shown in Figure 2(e).

(a)Input image (b) Face detected image

Normalized cuts segments of the input image is shown below:



(c) Normalized cuts segments

(d) Boundary detected image (e) Detected torso

Fig. 2: (a-e). Simulation results

VII. CONCLUSION

In this methodology, the torso portion of the human body is detected from static images by effectively combining the face detection, normalized cuts and boundary detection algorithms. Compared to previously existing torso detection approaches this method detects the torso portion of persons in images having various poses.

REFERENCES

[1] Huchan Lu, Guoliang Fang, Xinqing Shao and Xuelong Li, "Segmenting Human From Photo Images Based On a Coarse-to- Fine scheme", IEEE Transactions on Systems Man And Cybernetics, vol.42, no.3, june 2012
[2] G. Hua, M. H.Yang and Y.Wu, "Learning to estimate human pose with data driven belief propogation",in proceedings of CVPR 2005.
[3] G. Mori, X. Ren, A. Efros and J. Malik, "Recovering Human Body Configurations using pair-wise constraints between parts," in proceedings of CVPR , pp.824-832,2005.
[4] G. Mori, X. Ren, A. Efros and J. Malik, "Recovering Human Body Configurations:Combining Segmentation and Recognition," in proceedings of CVPR, pp.326-333, 2005.
[5] Z. Hu, X. Lin, and H. Yan, "Torso detection in static images," in Proceedings of IEEE Int. Conf. Signal Process., vol.3,2000.
[6] Paul Viola and Michael Jones, "Rapid Object Detection using a Boosted cascade of simple features., IEEE conference on computer vision and pattern recognition,2001.
[7] J. Shi and J.Malik, "Normalized cuts and image segmentation," IEEE Transactions on Pattern Analysis and MachineIntelligence, vol.22, no.8,Aug 2000.