The Human Image Segmentation Algorithm Based on Face Detection and Biased Normalized Cuts

Shaojun Qu(✉) and Qiaoliang Li

College of Mathematics and Computer Science, Hunan Normal University, Changsha 410081, China
{powerhope,lqlbox}@163.com

Abstract. Attributed to pose variation (frontal, profile, et.), the color and texture difference of clothes, the presence of noise, low contrast, uneven illumination and complex background. There are enormous difficultly in human image segmentation. In this paper, we propose an automatic human image segmentation method based on the face detection and biased normalized cuts. First, we use face detection algorithm to detect human faces, and get facial contours. Then we establish object seeds estimation model based on the position of the detected face, and get the object seeds. Using these seeds, we use biased normalized cuts algorithm to segment the image. Finally, we perform region merging based on the previous seeds and segmentation results, and the image is divided into two parts (object and background). We implement a large amount of experiments over a public segmentation database of Berkeley etc. Experiments show that our method can segment different types of human image and obtain satisfactory results. Compared with Grabcut method, our propose method can be obtained more accurate results in many images. Qualitative and quantitative experimental results demonstrate our method produces high quality segmentations and effectively improve the segmentation efficiency.

Keywords: Face detection · Human image segmentation · Biased normalized cuts · Seeds estimation model

1 Introduction

Human body segmentation in human images is a very important step in many computer vision tasks, such as image processing, video tracking, pose estimation, content-based image retrieval, pedestrian detection, action understanding, etc. However, to segment a human body in a human image is still a very challenging task because of segmentation is inherently ill-posed, the appearance and pose variation, the presence of noise, low contrast, and intensity inhomogeneity.

In the last decade the most popular approach to interacitve image segmentaiton in computer vision was graph cut. To avoid the minimum cut criteria favors cutting small sets of isolated nodes in the graph. Using the volume for

H. Zha et al. (Eds.): CCCV 2015, Part I, CCIS 546, pp. 134–143, 2015.
DOI: 10.1007/978-3-662-48558-3_14

the normalized weights. It aims at extracting the global impression of an image. The normalized cuts [1] criterion measures both the total dissimilarity between the different groups as well as the total similarity within the groups. Subhransu Maji [2] present a modification of "Normalized cuts" to incorporate priors which can be used for constrained image segmentation.

In this paper, we employ face detection and biaed normalized cuts to segment human body in static image. Different from the previous methods, our approach requires much less training data for face detection, and seeds estimation model is simple and effective. Moreover, our method is different from biased normalized cuts which we have better constraints and need to do a region merging after biaed normalized cuts segmentation. Also, our method does not require human interaction, and it is a fully automated segmentation method. Our segmentation results are more accurate and effective.

The rest of this paper is organized as follows. Section 2 discusses the most related work with ours. Section 3 describes the details of our proposed method. Analysis and experimental results in Section 4. Finally, Section 5 concludes the paper and propose some future work..

2 Related Work

2.1 Face Detection

Face detection is dominated by discriminatively-trained scanning window classifiers [3], most ubiquitous of which is the Viola Jones detector [4]. Zhu [5] model was based on a mixtures of trees with a shared pool of parts. They modeled every facial landmark as a part and used global mixtures to capture topological changes due to viewpoint. Their system was also trained discriminatively, but with much less training data, particularly when compared to commercial systems.

2.2 Human Image Segmentation

Ashwini T. Magar et. [6] divided human segmentation techniques to exemplar based, part based and other based. In exemplar based approach [7–9], an exemplar pool should be constructed first, and then, the test images was matched with the exemplars. Agarwal and Triggs [10] modeled the image window with a dense grid of local gradient orientation histograms to select similar human features, which made them capable of handling complex backgrounds. But the problem with these approaches were that in exemplar based approaches cannot always accurately segment the human body, because human poses are arbitrary and an exemplar pool cannot cover all the situations of poses and appearance variation.

In part based approaches, one can recover human body configurations by assembling set of candidate parts [11–13]. Drawback of these methods were that their performance depends on individual part detector and they were very hard to design a robust part detector.

In order to overcome drawback of both exemplar and part based approach. Some different techniques were developed. Mori [14] proposed a body model to estimate the human pose from static images based on superpixels. Shifeng Li et. [15] proposed a method to segment human body in static images by graph cuts based on two deformable models at two-scale superpixel. This method needed prior knowledge of the face. [16] presented a fully-automatic Spatio-Temporal GrabCut human segmentation methodology that combined tracking and segmentation. But the limitations of the method is that it depends on the initialization of the ST-GrabCut algorithm. However, these approaches typically have a large number of parameters, which leads to difficulty in calculating problem in high-dimensional space.

3 Our Algorithm Based on Face Detection and Biased Normalized Cuts

An overview of our method is show in Fig. 1. The whole segmentation methodology is detailed in algorithm 1. We combined face detection and biased normalized cuts. First, we performed face detection for input image(line 1-2).

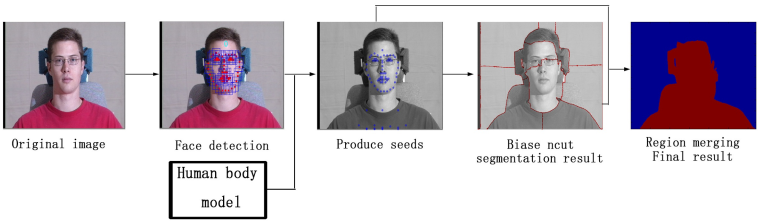

Fig. 1. Overview of our method.

Step 2(line 3). Compute seeds based on face detection results and human body model. The detailed in algorithm 2. Through statistical analysis propotional relationship of various parts of the body and the relationship between the size of human after we have carried out a lot of experiments. We established a human body model to choose seeds automatically in Fig. 2. According to recognized human faces, we deduce the coordinate of the rest of seeds. The rectangle(red) represents the recognized face region. It's width is w, height is h, Vertex coordinates of the upper left corner is $(minx, miny)$ and lower right corner of the vertex coordinates is $(maxx, maxy)$. We calculate the 23 seeds coordinates of other body parts based on the model. The seeds coordinates are $[(minx + w/2, miny - 15), (minx, maxy + h/2), (minx + w/2, maxy + h/2), (maxx, maxy + h/2), (minx, maxy + h), (minx + w/4, maxy + h), (minx + w/2, maxy+h), (maxx-w/4, maxy+h), (maxx, maxy+h), (minx-w/4, maxy+3*h/2), (maxx+w/4, maxy+3*h/2), (minx, maxy+2*h), (minx+w/4, maxy+$

Algorithm 1 Our algorithm based on face detection and biased normalized cuts(G, w, S_T, γ)

1: im =read input image.
2: Face detection for im.
3: Compute seeds S_T based on face detection results and human body model(Algorithm 2).
4: Construction graph $G = (V, E)$, compute edge weight, the similarity matrix w based on intervening contours, a correlation parameter $\gamma \in (-\infty, \lambda_2(G))$
5: $A_G(i,j) \leftarrow w(i,j), D_G(i,j) \leftarrow \sum_j w(i,j)$
6: $L_G \leftarrow D_G - A_G, \text{Ł}_G \leftarrow D_G^{-1/2} L_G D_G^{-1/2}$
7: Compute $u_1, u_2, \cdots, u_K$ the eigenvectors of Ł_G corresponding to the K smallest eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_k$
8: $w_i \leftarrow \frac{u_i^T D_G S_T}{\lambda_i - \gamma}$, for $i = 2, \cdots K$
9: Obtain the biased normalized cuts, $x* \propto \sum_{i=2}^{K} w_i u_i$
10: Region merging.
11: Output results.

$2*h), (minx + w/2, maxy + 2*h), (maxx - w/4, maxy + 2*h), (maxx, maxy + 2*h), (minx, maxy + 3*h), (minx + w/2, maxy + 3*h), (maxx, maxy + 3*h), (minx, maxy + 4*h), (maxx, maxy + 4*h), (minx, maxy + 5*h), (maxx, maxy + 5*h),]$. Finally, we detect seeds that falls outside the human body and the image, and then correcting the seeds. Green dots represent inferred seeds by algorithm 2.

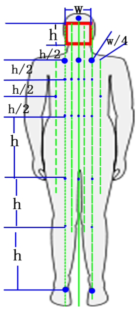

Fig. 2. Human body model.

Step 3(line 4-9). Step2's seeds as biased normalized cuts constraints condition. We use biased normalized cuts algorithm to segment the image, and obtain the initial segmentation result.

Step 4(line 10). Region merging. First, we merge the region where have a seed as the object. Then from top to down, from left to right continue scan image, if there are more than two object regions around the undetermined regions, we merged this region into object. Finally, we considered the undetermined regions

Algorithm 2 Automatic produce seeds

```
1: if  the number of identify the face detection  == 1 then
2:     According to face detection results and human body model to compute seeds.
3:     Detect seeds that falls outside the image, and correcting the seeds.
4: else
5:     for each i ∈ the number of identify the face region do
6:         Compute seeds for region i based on face detection results and human body
           model.
7:         Detect seeds that falls outside the image, and correcting the seeds.
8:     end for
9: end if
```

as a background region, we can obtain segmentation results that include two parts(object and background).

4 Analysis and Experiments

On the basis of normalized cuts, biased normalized cuts introduced the constraint condition, and convert image segmentation problem became a constrained problem. Our approach strengthen and optimize the constraints. Moreover, different from biased normalized cuts which we need to do a region merging after segmentation, we can obtain segmentation results that include two parts(object and background). Our segmentation results are more accurate and effective.

We evaluate the proposed method on three datasets. The first one is the Berkeley segmentation dataset [17] which contains 78 images of only one person. The second one is the Grabcut dataset [18] which contains 16 images. The last one is Ramanan's annotated dataset [5]. In order to quantitatively evaluate our method, we chose 31 representative of images from which these datasets. These images cover front, side, half-length, whole body, complex background, all kinds of clothes, light conditions, as shown in Fig. 3.

The evaluation setup is as follows. Manually segmentations are used to construct ground-truth segmentation, this produces a binary segmentation that will be used as ground truth, as shown in Fig. 4.

First, we assess quality of seed selection by our human body model. In addition to the two images in Fig. 7 , there are some error seeds, other pictures can get right seeds. Image 189011.jpg identified two face, which one is wrong. We found that the accuracy of automatically obtained seeds is closely related to the accuracy of face recognition. Results indicates that our human body model is a good seed selection method for image segmentation.

Select parameter. It is necessary to set the number of segment block(*nbSegments*) since our method based on biased normalized cuts, In our method, we chose the *nbSegments* to be 12 or 15 or 20 or 30 or 50. We then choose a relatively desirable results, as shown in Table 1. Our method can accurately segment human body on the dataset. As shown in Fig. 5.

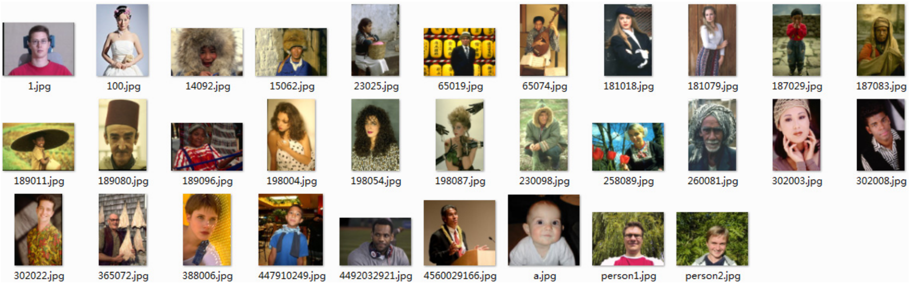

Fig. 3. Original image.

Fig. 4. Ground truth by human segmentations.

Fig. 5. Our method segmentation results.

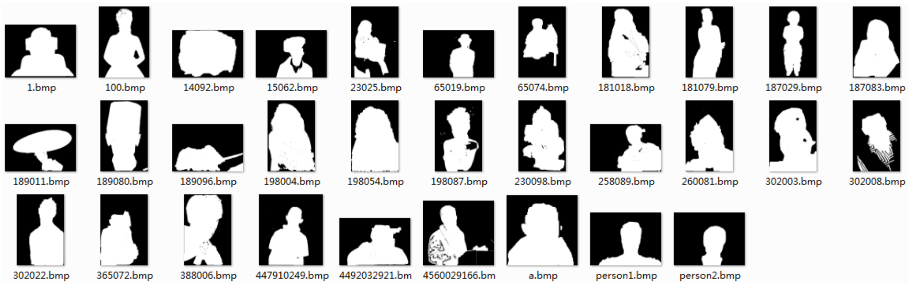

Fig. 6. Grabcut segmentation results.

Fig. 7. Exist error seeds in images.

Table 1. Select the number of segmentation blocks

nbSegments	The optimal number of results	nbSegments	The optimal number of results
12	5	15	10
20	10	30	5
50	1		

Comparison with Other Methods. We apply the proposed method to gether with the existing Grabcut method. GrabCut is a way to perform 2D segmentation in an image that is very user friendly. The user only need to input the a very rough segmentation between foreground and background. The initial information given about the foreground and the background are given by the user as a rectangular selection around the object of interest. Pixels outside this selection are treated as known background and the pixels inside are marked as unknown.

In experiment, we selected the object use polygon, instead of rectangle, and the object bounding box is comparatively precise. Parameters $k = 6$ or 9, $\beta = 0.3$. The Grabcut optimum results as show in Fig. 6.

Accuracy of Our Method. To test the accuracy of our method and Grabcut, we use Fig. 4 as the ground truth, and use Jaccard similarity coefficient [19] and a Modified Hausdorff distance [20] to evaluate the accuracy.

The Jaccard coefficient measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}. \tag{1}$$

If A and B are both empty, we define $J(A,B) = 1$. Clearly, $0 \leq J(A,B) \leq 1$. The more similar A and B, the more the value is larger.

Fig. 8 reports our approach and grabcut compared with the ground truth. The dots represent our methods and the asterisks represent grabcut. In most of images, our method obtain satisfactory results.

The Hausdorff Distance. Given two finite point sets $A = \{a_1, \cdots, a_p\}$, $B = \{b_1, \cdots, b_p\}$, the Hausdorff distance [21] is defined as

$$H(A,B) = max(h(A,B), h(B,A)). \tag{2}$$

where $h(A,B) = \max_{a \in A} \min_{b \in B} ||a - b||$. and $||.||$ is some underlying norm on the points of A and B(e.g., the L_2 or Euclidean norm). A Modified Hausdorff

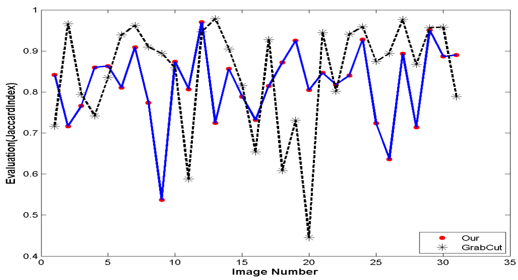

Fig. 8. Comparison between grabcut and our method with jaccardIndex.

distance [20] introduces 24 possible distance measures based on the Hausdorff distance between two point sets. Our method and Grabcut compared with the ground truth, with the results as show in Fig. 9.

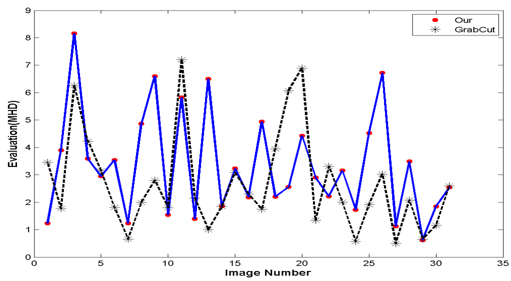

Fig. 9. Comparison between grabcut and our method with MHD.

Finally, Given two finite point sets A and B , the mean error rtate(ME) is defined as

$$ME = \frac{1 - p_1 - p_4}{r \times \ c}. \tag{3}$$

p_1 represents the number of A and B correspond to the position of the pixel label are background. p_4 represents the number of A and B correspond to the position of the pixel label are object. Results is shown in Fig. 10. The results show that our method is superior to grabcut in a lot of images.

Comparison shows that more than a third of the results are better than GrabCut. There are several results are the same. The grabcut is better than our method in other results. Experiments shows that Grabcut is sensitive to body pose. If the pose varies drastically, the results of Grabcut are imprecise. So, in our experiment, we selected the object use polygon, instead of rectangle, and the object bounding box is comparatively precise. From these examples, we can

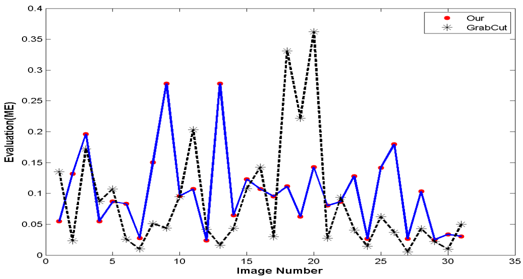

Fig. 10. Comparison between grabcut and our method with Mean error.

see that in most cases, our method can segment the human body accurately. Moreover, our method automatic select seeds and automated segmentation can reduce human interaction time, greatly improving the efficiency of segmentation.

5 Conclusions and Future Work

We proposed a framework for human body segmentation using face detection and biased normalized cuts, and present a simple and effective method for computing seeds for biased normalized cuts. First, we used face detection technology to identify face of the human, and used a human body model to estimate the lower body by recognized human face, then using a part of face and body as seeds, we use biased normalized cuts algorithm to segment the image. Finally, we conducted region merging, and the image is divided into two parts(object and background). Our algorithm could segment the whole human from the image. Experiment demonstrated that our method could reduce the time of human interaction and the efficiency of segmentation. The main limitation of our approach is that it depends on a correct detection of the person and his/her face.

As a future work, the algorithm could be extended in order to segment more than one person present in the images, since our current method only segments one subject in the images.

Acknowledgments. The work was supported by the National Natural Science Foundation of China(NSFC), NO. 11471002 and Hunan Provincial Science and Technology Plan, NO. 2013FJ4052.

References

1. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**, 888–905 (1997)
2. Maji, S., Vishnoi, N.K., Malik, J.: Biased normalized cuts. In: CVPR, pp. 2057–2064. IEEE Computer Society, Colorado Springs (2011)

3. Heisele, B., Serre, T., Poggio, T.: A component-based framework for face detection and identification. International Journal of Computer Vision **74**, 167–181 (2007)
4. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision **57**, 137–154 (2004)
5. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: CVPR, pp. 2879–2886. IEEE Computer Society, Providence (2012)
6. Magar, A.T., Shinde, J.V.: A survey of techniques for human segmentation from static images. In: International Association of Scientific Innovation and Research, pp. 66–75. United State (2014)
7. Kohli, P., Rihan, J., Bray, M., Torr, P.H.: Simultaneous segmentation and pose estimation of humans using dynamic graph cuts. International Journal of Computer Vision **79**, 285C–298 (2008)
8. Lin, Z., Davis, L., Doermann, D., DeMenthon, D.: Hierarchical part-template matching for human detection and segmentation. In: ICCV, pp. 1–8. IEEE Press, Brazil (2007)
9. Kumar, M.P., Torr, P., Zisserman, A.: Objcut: Efficient segmentation using top-down and bottom-up cues. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**, 530–545 (2010)
10. Agarwal, A., Triggs, B.: A local basis representation for estimating human pose from cluttered images. In: Narayanan, P.J., Nayar, S.K., Shum, H.-Y. (eds.) ACCV 2006. LNCS, vol. 3851, pp. 50–59. Springer, Heidelberg (2006)
11. Mori, G., Ren, X., Efros, A., Malik, J.: Recovering human body configurations: combining segmentation and recognition. In: CVPR, pp. 326–333. IEEE Computer Society, Washington (2004)
12. Ren, X., Berg, A., Malik, J.: Recovering human body configurations using pairwise constraints between parts. In: ICCV, pp. 824–831. IEEE Computer Society, Beijing (2005)
13. Hua, G., Yang, M.H., Wu, Y.: Learning to estimate human pose with data driven belief propagation. In: CVPR, pp. 747–754. IEEE Computer Society, San Diego (2005)
14. Mori, G.: Guiding model search using segmentation. In: ICCV, pp. 1417–1423. IEEE Computer Society, Beijing (2005)
15. Li, S., Lu, H.C., Ruan, X., Chen, Y.W.: Human body segmentation based on deformable models and two-scale superpixel. Pattern Analysis and Applications **15**, 399–413 (2012)
16. Hernndez-Vela, A., Reyes, M., Ponce, V., Escalera, S.: Grabcut-based human segmentation in video sequences. Sensors **12**, 15376–15393 (2012)
17. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV, pp. 416–423. IEEE Computer Society, Vancouver (2001)
18. Rother, C., Kolmogorov, V., Blake, A.: Grabcut - interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. **23**, 309–314 (2004)
19. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall, Upper Saddle River (1988)
20. Dubuisson, M.P., Jain, A.K.: A modified Hausdorff distance for object matching. In: CVPR. IEEE Computer Society, Seattle (1994)
21. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing Images Using the Hausdorff Distance. IEEE Transactions on Pattern Analysis and Machine Intelligence **15**, 850–863 (1993)