Level Set Based Automatic Human Body Segmentation

Muhammad Hameed Siddiqi, Phan Tran Ho Truc, Sungyoung Lee, Young-Koo Lee
Ubiquitous Computing Lab, Department of Computer Engineering, College of Electronics & Information, Kyung Hee University (Global Campus), Korea.
{siddiqi, pthtruc, sylee}@oslab.khu.ac.kr, yklee@khu.ac.kr

Abstract: *The accuracy of the video-based AR depends significantly on the performance of human body segmentation. Automatically human body segmentation is one of the most important and challenging issue in the field of computer vision and pattern recognition for u-Life care. Existing methods often involve modeling of the human body and/or the background, which normally requires extensive amount of training data and cannot efficiently handle changes over time. Recently, active contours have been emerging as an effective segmentation technique in still images. In this paper, we propose to adapt an active contour model that segments the dynamic images automatically, which is robust to illumination and clothing changes, typical issues in practical Activity Recognition systems. The experimental results of our approach show good segmentation results.*

Keywords*: *Active contour, Segmentation, Computer vision.

1. INTRODUCTION

Automatically human body segmentation is one of the most important and challenging issue in the field of computer vision and pattern recognition for u-Life care. One of the main targeting services of u-Life care is to enable people to live independently longer through the early detection and prevention of chronic disease and disabilities. Computer vision, emplaced wireless sensor networks (WSN), and body networks are emerging technologies that promise to significantly enhance medical care for seniors living at home in assisted living facilities. With these technologies, we can collect video, physiological, and environmental data, identify individuals' activities of daily living (ADL), and act for improved daily medical care as well as real-time reaction to medical emergencies. Overall, projected benefits include greater independence for the elderly, lower medical costs through reduction in hospital and emergency room visits, improved health, and via longitudinal studies, increased understanding of the causes of diseases and the efficacy of their treatment. Activity recognition can be applied to many applications which can roughly be grouped into four general domains, namely smart surveillance, virtual reality, advanced user interfaces, and motion analysis.

The accuracy of the video-based activity recognition depends on the accuracy of the human body segmentation. To segment the human body automatically, is one of the main and important issue in the field of computer vision and pattern recognition. Many existing works have the problems of segmenting the human body automatically from the background [1]. In our previous work [2], we also have the problem of segmenting the human body automatically from the background; we just subtracted the empty frame from activity frame to segment the human body as shown in Fig.1.

Fig.1- Segmentation based on subtracting empty frame from the activity frame

The author of [3] developed an algorithm for image subtraction for real time moving object. In their method they have obtained the motion mask by applying background subtraction and consecutive frame differencing, and then finally updated the background by using noise reduction operator that facilitate the result of moving object extraction. The limitation of this work is that it cannot work for static activities (like bending, jacking, hand waving etc.), because in these types of activities the human body have the same position. If we will subtract the consecutive frames from each other, then it will lose a lot of information. In [4], the author presented a method for real time background segmentation. In his method he represented each pixel in the frame by the group of clusters and the clusters are ordered accordingly, due to which the background has been modeled and adapted to deal with background and lighting variations. So by this way the incoming pixels are matched with the clusters to classify that weather the corresponding pixels belongs to the part of background or not. The limitation of this work is that, it cannot provide better result for human activity recognition, because the output of this work is like the output of edge detection which loses a lot of depth information that is related to human activity. The objective of this paper was to cover the limitations of the above methods and was to develop a new algorithm that can easily segment foreground (human body) from the background.

2. METHODOLOGY

The accuracy of the video-based AR depends significantly on the performance of human body segmentation. Object segmentation is the process of separating the objects of interest (human body) from the rest of the image (the background). Methods for object segmentation are often applied as the first step in many systems and therefore a crucial process. In the field of image segmentation, since it was first introduced by [5], Active Contour (AC) model has attracted much attention. Recently, Chan and Vese (CV) proposed in [6] a novel form of AC based on the Mumford and Shah functional for segmentation and the level set framework. Unlike other AC models which rely much on the gradient of the image as the stopping term and thus have unsatisfactory performance in noisy images, the CV AC model does not use the edge information but utilizes the difference between the regions inside and outside of the curve, making itself one of the most robust and thus widely used techniques for image segmentation. Its energy functional is defined by

$$F(C)=\int_{in(C)}\left|I(\mathbf{x})-c_{in}\right|^2 d\mathbf{x}+\int_{out(C)}\left|I(\mathbf{x})-c_{out}\right|^2 d\mathbf{x} \tag{1}$$

where $\mathbf{x}\in\Omega$ (the image plane) $\subset R^2$, $I:\Omega\rightarrow\mathcal{Z}$ is a certain image feature such as intensity, color, or texture, etc., and c_{in} and c_{out} are respectively the mean values of image feature inside $[in(C)]$ and outside $[out(C)]$ the curve C, which represents for the boundary between two separate segments. Considering image segmentation as a clustering problem, we can see that this model forms two segments (clusters) such that the differences within every segment are minimized. However, the global minimum of the above energy functional does not always guarantee the desirable results, especially when a segment is highly inhomogeneous, e.g., human body, as can be seen in Fig. 2(b). The unsatisfactory result of the CV AC in this case is due to the fact that it is trying to minimize the dissimilarity within each segment but does not take into account the distance between different segments. Our methodology is to incorporate an evolving term based on the Bhattacharyya distance to the CV energy functional such that not only the differences within each region are minimized but the distance between the two regions is maximized as well. The proposed energy functional is

$$E_0(C)=\beta F(C)+(1-\beta)B(C) \tag{2}$$

where $\beta\in[0,1]$, $B(C)\equiv B=\int_{\mathcal{Z}}\sqrt{p_{in}(z)p_{out}(z)}dz$ the Bhattacharyya coefficient [7]

$$p_{in}(z)=\frac{\int_\Omega \delta(z-I(\mathbf{x}))H(-\phi(\mathbf{x}))d\mathbf{x}}{\int_\Omega H(-\phi(\mathbf{x}))d\mathbf{x}} \tag{3}$$

$$p_{out}(z)=\frac{\int_\Omega \delta(z-I(\mathbf{x}))H(\phi(\mathbf{x}))d\mathbf{x}}{\int_\Omega H(\phi(\mathbf{x}))d\mathbf{x}} \tag{4}$$

$\phi:\Omega\rightarrow R$ the level set function, and $H(\cdot)$ and $\delta(\cdot)\triangleq H'(\cdot)$ respectively the Heaviside and the Dirac functions [8]. Note that the Bhattacharyya distance is defined by $[-\log B(C)]$ and the maximization of this distance is equivalent to the minimization of $B(C)$. Note also that to be comparable to the $F(C)$ term, in our implementation, $B(C)$ is multiplied by the area of the image because its value is always within the interval $[0,1]$ whereas $F(C)$ is calculated based on the integral over the image plane. In general, we can regularize the solution by constraining the length of the curve and the area of the region inside it. Therefore, the energy functional is defined by

$$E(C)=\gamma\int_\Omega|\nabla H(\phi(\mathbf{x}))|d\mathbf{x}+\eta\int_\Omega H(-\phi(\mathbf{x}))d\mathbf{x}+\beta F(C)+(1-\beta)B(C) \tag{5}$$

where $\gamma\geq 0$ and $\eta\geq 0$ are constants.

The intuition behind the proposed energy functional is that we seek for a curve which 1) is regular (the first two terms) and 2) partitions the image into two regions such that the differences within each region are minimized (i.e., the $F(C)$ term) and the distance between the two regions is maximized (i.e., the $B(C)$ term).

The level set implementation for the energy functional in (5) can be derived as:

$$\frac{\partial\phi}{\partial t}=|\nabla\phi|\left\{\begin{array}{l}\gamma\kappa+\eta+\beta\left[(I-c_{in})^2-(I-c_{out})^2\right]\\ -(1-\beta)\left[\frac{B}{2}\left(\frac{1}{A_{in}}-\frac{1}{A_{out}}\right)+\frac{1}{2}\int_{\mathcal{Z}}\delta(z-I)\left(\frac{1}{A_{out}}\sqrt{\frac{p_{in}}{p_{out}}}-\frac{1}{A_{in}}\sqrt{\frac{p_{out}}{p_{in}}}\right)dz\right]\end{array}\right\} \tag{6}$$

where A_{in} and A_{out} are respectively the areas inside and outside the curve C.

As a result, the proposed model can overcome the CV AC's limitation in segmenting inhomogeneous objects as shown in Fig. 2(c), yielding the body detector more robust to illumination changes and clothing.

(a) (b) (c)

Fig.2- Sample segmentation of inhomogeneous body-shape object using active contours. (a) Initial contour, (b) result of CV AC, and (c) result of our approach. The CV AC fails to capture the whole body whereas our approach can.

3. RESULTS AND DISCUSSION

In order to evaluate the proposed algorithm, we used a publicly available dataset [9]. In this dataset, video clips of nine activities were recorded, namely "bend", "jack" (jumping-jack), "jump" (jumping forward on two legs), "run", "side" (gallopsideways), "skip", "walk", "wave1" (wave-one-hand), and "wave2" (wave-two-hands). Each activity was performed by nine different people. The frame size is 144 x 180. The proposed segmentation approach aims for human body extractions automatically that will be used for human activity recognition. Human body segmentation in video data is done frame-based, which means that the active contour evolution in a certain frame is performed independently of other frames. The only utilized information is the final contour obtained in the previous frame which will be used to determine the initial position of the active contour in the current frame. The process is outlined as follows.

- The initial contour is selected as an ellipse with major axis along *y*-axis and of length 50 and minor axis along *x*-axis and of length 20. This initial shape will be the same for all frames in this experiment and other similar experiments in this paper; only its center's location varies.

- In each video, the first frame is segmented using manual initialization such that the initial contour is close to the object.
- From the second frame, the position of the initial contour's center in the current frame is the mass center (mean value) of points along the final contour in the previous frame. For example, suppose that along the final contour of frame $k(\geq 1)$, there are N points $(x_i^{(k)}, y_i^{(k)}), i = 1...N$. Then, the center $(c_x^{(k+1)}, c_y^{(k+1)})$ of the initial contour in frame (*K+1*) is calculated as

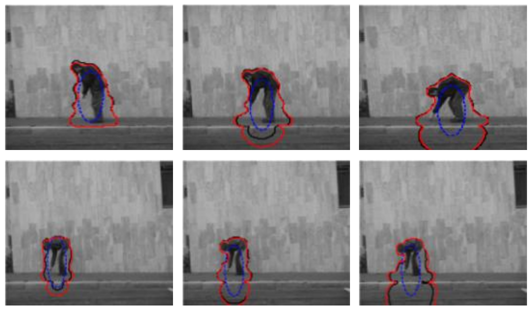
$$c_x^{(k+1)} = \frac{1}{N}\sum_{i=1}^{N} x_i^{(k+1)}; \quad c_y^{(k+1)} = \frac{1}{N}\sum_{i=1}^{N} y_i^{(k+1)}$$

Some sample segmentation results on images of different activity videos are shown in Fig.3. We can see that the proposed model with the above-described scheme works well with "dynamic" activities such as "skip", or "run", or "walk". In our previous work [10], the proposed technique works well on "static" activities such as "bend" or "wave". However, for "dynamic" activities such as "run", "skip", or "walk", it fails to capture the whole body correctly. In this paper we have presented the qualitative segmentation results and the quantitative evaluation will be performed indirectly via the accuracy of the activity recognition system that is the future work of this paper. The overall segmentation results of the proposed algorithm on different video activities are given in Fig.3.

4. CONCLUSION

This paper has presented an active contour model for human body segmentation from video data. Like other AC models, when applied to video data where the background environment is much more arbitrary compared to that of the medical imagery, it requires a less relaxed initialization scheme, i.e., the initial contour should be close to the object in order to correctly converge. A straightforward way is to use the segmentation result in the previous frame. Specifically, the mass center of points along the final contour (corresponding to the object boundary) in the previous frame is used as the center of the initial contour in the current frame, and also used as the center of the of the initial contour in the current frame As a result, the proposed AC model with this initialization scheme can correctly automatically segment the human body in both "static" and "dynamic" activity videos.

The proposed algorithm works well for both still and dynamic activities, but for some static activities like bending activity, it cannot segment the human body well, because in this algorithm we move the initial contour up and down (means x and y direction), due to which the final contour expand, which move downward. This is the limitation of the proposed model, the output is given as:



In future we will try to modify the proposed technique to solve the above problem and will try to segment the human body from every type of still and dynamic activities.

ACKNOWLEDGMENT

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency)" (NIPA-2010-(C1090-1021-0003)).

5. REFERENCES

[1] M.Z. Uddin, J.J. Lee, and T.-S. Kim, Shape-Based Human Activity Recognition Using Independent Component Analysis and Hidden Markov Model, *Proc. of 21st International Conference on Industrial, Engineering, and other Applications of Applied Intelligent Systems*, pp.245-254, Springer-Verlag Berlin Heidelberg, 2008.

[2] M.H. Siddiqi, M. Fahim, S.Y. Lee, and Y.K. Lee, Human Activity Recognition Based on Morphological Dilation followed by Watershed Transformation Method, *Proc. of International Conference on Electronics and Information Engineering (ICEIE)*, V2, pp. 433, 2010.

[3] S.M. Desa, and Q.A. Salih, "Image Subtraction for Real Time Moving Object Extraction," *Proc. of the International Conference on Computer Graphics, Imaging and Visualization (CGIV)*, pp. 41–45, 2004.

[4] D. Butler, S. Sridharan, and V.M.B. Jr, "Real Time Adaptive Background Segmentation", *Proc. of the International Conference on Multimedia and Expo (ICME)*, Vol 3, pp. 341 – 344, 2003.

[5] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Int. J. Comput. Vis.*, vol. 1, pp. 321-31, 1988.

[6] M. Yamamoto, H. Mitomi, F. Fujiwara, T. Sato, Bayesian classification of task-oriented actions based on stochastic context-free grammar, in: *Int. Conf. Automatic Face and Gesture Recognition*, Southampton, UK, April 10–12, 2006.

[7] T. Kailath, The divergence and Bhattacharyya distance measures in signal selection, *IEEE Trans. Commun. Technol*. vol. 15, pp. 52–60, 1967.

[8] T. Chan and L. Vese, Active contours without edges, *IEEE Trans. Image Proc*. 10 (2001) 266-277.

[9] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, Actions as Space-Time Shapes, IEEE Trans. PAMI, vol. 29, no. 12, pp. 2247-53, 2007.

[10] M.H. Siddiqi, M. Fahim, P.T.H. Truc, Y.K. Lee, and S.Y. Lee. Active Contour Based Human Body Segmentation with Applications in u-Life Care. Proc. Of the 7th International Conference on Ubiquitous Health Care (u-Healthcare'10), Jeju, Korea.

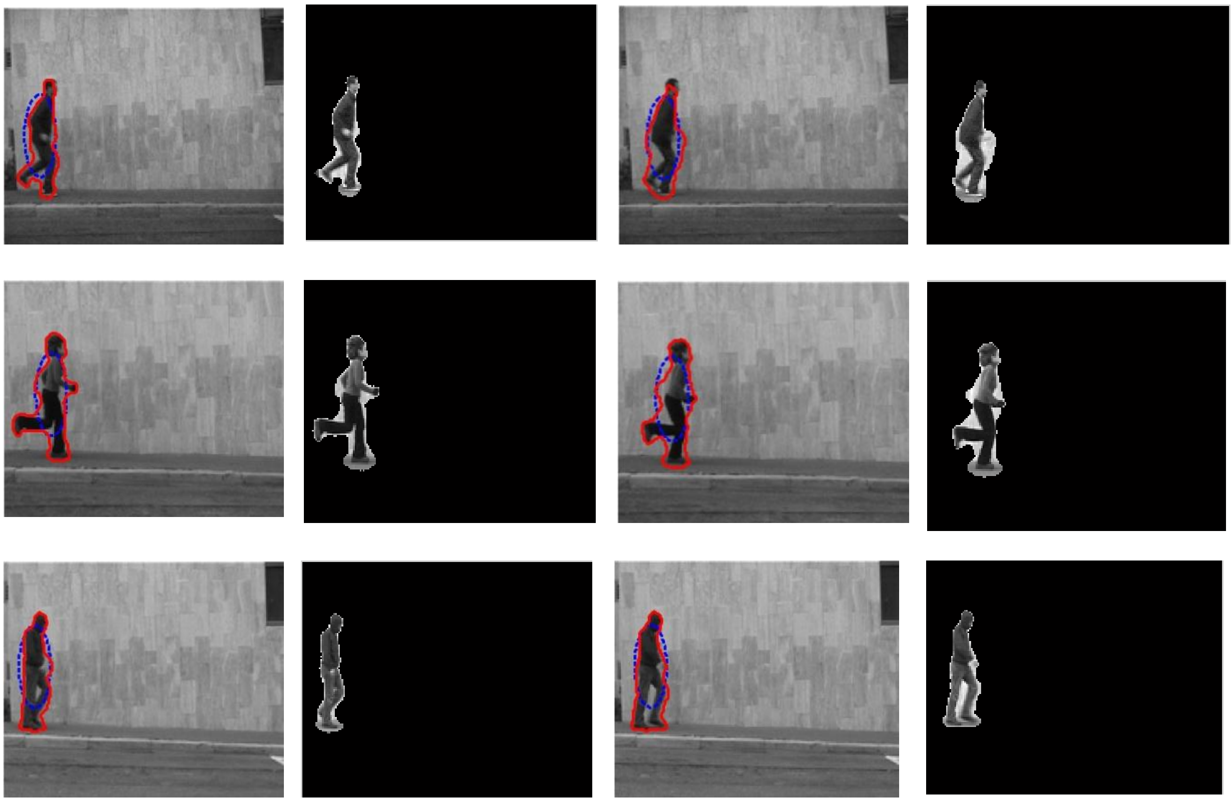


Fig.3- Segmentation results of different types of activities (skip, run, and walk). The blue circle is the initial contour and the red one is the final contour for the current frame.