Active Contours Level Set Based Still Human Body Segmentation from Depth Images For Video-based Activity Recognition

Muhammad Hameed Siddiqi[1], Adil Mehmood Khan[2] and Seok-Won Lee[2],*

[1] Department of Computer Engineering, Kyung Hee University, 1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, Rep. of Korea
[e-mail: siddiqi@oslab.khu.ac.kr]
[2] Division of Information and Computer Engineering, Ajou University, San 5, Woncheon-dong, Yeongtong-Gu, Suwon, Rep. of Korea
[e-mail: {amtareen, leesw}@ajou.ac.kr]
Corresponding author: Seok-Won Lee

Abstract

Context-awareness is an essential part of ubiquitous computing, and over the past decade video based activity recognition (VAR) has emerged as an important component to identify user's context for automatic service delivery in context-aware applications. The accuracy of VAR significantly depends on the performance of the employed human body segmentation algorithm. Previous human body segmentation algorithms often engage modeling of the human body that normally requires bulky amount of training data and cannot competently handle changes over time. Recently, active contours have emerged as a successful segmentation technique in still images. In this paper, an active contour model with the integration of Chan Vese (CV) energy and Bhattacharya distance functions are adapted for automatic human body segmentation using depth cameras for VAR. The proposed technique not only outperforms existing segmentation methods in normal scenarios but it is also more robust to noise. Moreover, it is unsupervised, i.e., no prior human body model is needed. The performance of the proposed segmentation technique is compared against conventional CV Active Contour (AC) model using a depth-camera and obtained much better performance over it.

Keywords: Human body segmentation, active contour, Bhattacharyya distance, activity recognition, depth camera

1. Introduction

At the heart of ubiquitous and pervasive computing lies context-awareness. It is the context of the user that assists a ubiquitous application in automatic service delivery of the right kind, such as making an emergency call to the hospital in case of a fall incident of an elderly home-living patient or turning the lights off when a user has left his office.

Activity recognition is one of the key components in identifying the context of a user for providing the services based on the type of ubiquitous applications. For example, in case of ubiquitous healthcare applications, recognition of everyday activities could enable such systems to watch and learn any changes in daily behavior of an elderly person that might be the indicators of developing physical or mental medical conditions. Also, it could help to determine the level of independence of elderly people, to understand side effects of medication, and to encourage medication adherence [1, 2].

Video-based human activity recognition means automatic recognition of physical activities of human by a computer using video cameras. The accuracy of such systems depends significantly on the performance of human body segmentation. The reason why human body segmentation is so critical in such systems is because it defines the image area necessary and sufficient for the follow-up modules like feature extraction. In other words, it defines the body shapes used for feature extraction, which directly determines the recognition accuracy. Research has shown that trying to locate a human body (object and human body are used interchangeably at some places) contour purely by running a low level image processing task such as canny edge detection is not particularly successful technique because mostly the edges are not continuous and serious edges can be present because of noise [3].

The objective of this research was to develop an unsupervised automatic AC model for human body segmentation from depth images. The proposed AC model is the combination of the two energy functions Chan Vese (CV) and Bhattacharya Distance that not only minimizes the dissimilarities within the object but also maximizes the distance between the object and the background. Compared to the conventional CV AC model for static depth image segmentation, the proposed model is not only more accurate in normal scenarios but it is also more robust to noise variations. However, like other AC models, when applied to video data where the background environment is much more arbitrary, it requires a less relaxed initialization scheme, i.e., the initial contour should be close to the object in order to correctly converge. When compared against CV AC model for numerous physical activities, such as body bending, place humping, one hand waving, two hand waving, clapping and boxing, using a depth-camera, the proposed model provided much better results.

The rest of the paper is organized as follows. Section 2 discusses some related work about human body segmentation. The proposed approach is described in detail in section 3. Then the segmentation results and discussion is presented in comparison with that of the conventional CV AC model and with some state-of-the-art methods in section 4. Finally, the paper is concluded with some future directions in section 5.

2. Related Work

Human body segmentation is a process that is used to extract the human body shape from video frames and generate corresponding images. Mostly, the segmentation methods can be categorized into region or boundary-based methods [4]. As for the region-based category, several methods have been implemented, such as splitting and merging [5], watershed and motion-based segmentations [6, 7], and a primitive method for video object extraction [8]. However, the abovementioned approaches produce artifacts due to occlusion.

Currently, no one can point out which is the most advantageous solution due to different constrictions [9]. The author of [10] used a method based on similarity close measure to classify the belonging of the pixel followed by region growing to get the objects. But for segmentation, a set of markers was required, and it was also very hard for this method to discriminate which part should be segmented if there is an unknown image. To solve the abovementioned problem, active contour models were introduced by [11] that tried to move the contour towards the object of interest.

Some boundary-based methods were developed by [12 – 14] that mostly started with initial curve and used the gradient information to locate the object boundaries. All the aforementioned approaches had limited accuracy, because all the approaches were completely dependent on the initial curve that was explicitly defined and initialized by the user manually, which must be near the boundary of the desired object to be segmented. Also, if more than two objects needed to be segmented, two different contours needed to be initialized, and the user had to decide which object needed to be segmented by initializing it near the object.

The authors of [15 – 17] segmented the human body just by subtracting the empty frame from the RGB frame captured by video camera, and then generated the consequent binary silhouettes. However, these techniques were not applicable in real world environment, because in some situation, if there is no empty frame, it makes it very hard to segment the human body by employing these techniques. Due to this limitation, these methods are known as heuristic techniques.

Moreover, even though binary silhouettes have been extensively applied to symbolize a variety of body configuration, they sometimes generate ambiguities by representing the same silhouette for different postures from different activities. For example, if a person performs some hand movement normal to the video camera, different postures can correspond to the same silhouettes because of its binary level flat pixel intensity distribution [18]. In such cases, binary silhouettes do not seem to be the preferred choice for distinguishing these different postures from different activities.

Fig. 1 shows the RGB, binary and depth frames of hands-up-down, clapping, and boxing activities respectively. It is obvious that the binary silhouettes are a poor choice to separate these different postures. Unlike binary silhouettes, where there is a flat binary distribution of the pixels [18], in depth silhouettes body pixels are distributed based on the distance to the camera, making them a superior choice over binary silhouettes. The depth silhouettes can be obtained by using the infrared sensor-based depth camera or disparity calculation of the pixels in the stereo RGB images captured using a stereo camera [16].

Just like RGB-camera based techniques, such as [15 – 17], some of the existing depth-based works, such as [18 – 20], segmented the human body from depth just by subtracting the empty frame from depth video frame. Due to that reason these methods were also known as heuristic techniques.

Another reason for choosing the depth-camera over RGB-cameras for our work is the fact that ubiquitous application that employ video technologies raise privacy concerns since it can lead to situations where subjects may not know that their private information is being shared and thus become exposed to a threat [23]. Unlike RGB-cameras, depth-cameras only capture the depth information and do not reveal the identity of the subject or other sensitive information, which makes them a superior choice over RGB-cameras.
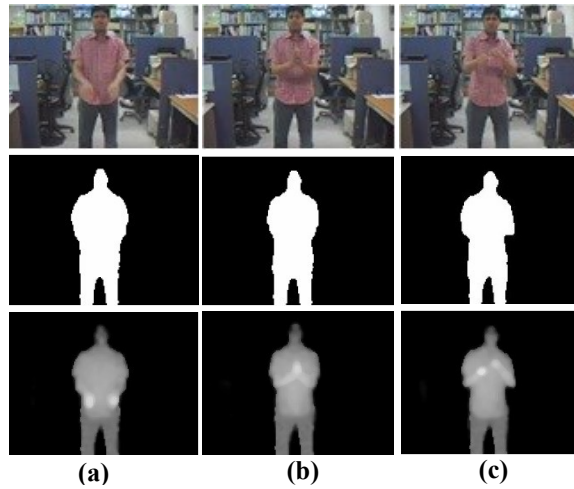
(a)   **(b)**   **(c)**

Fig. 1. Different types of human activities. (a) shows both hands up and down, (b) represents hand clapping and (c) indicates boxing activities respectively [18]. It is obvious from the figure that the binary silhouettes cannot provide good features through which we can discriminate such types of activities, but on the other hand depth frames provides best features for recognizing these activities easily.

3. Materials and Methods

The proposed segmentation model performs two parallel tasks to address two problems: 1) minimizing the dissimilarities within an object, and 2) maximizing the distance between the two regions (in our case human body and background).

AC model has attracted much attention in the field of image segmentation that was first introduced by Kass et al., [21]. Recently, Chan and Vese proposed in [11] a novel form of AC model for segmentation based on 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 such as human body segmentation. Its energy function is defined by

$$F(C) = \int_{in(c)} \left| I(x) - C_{in} \right|^2 dx + \int_{out(c)} \left| I(x) - C_{out} \right|^2 dx \qquad (1)$$

where $x \in \Omega$ (the image plane) $\subset R^2$, $I:\Omega \rightarrow Z$ is a certain image feature such as intensity, color, or texture, and C_{in} and C_{out} are respectively the mean values of image feature inside *[in(C)]* and outside *[out(C)]* the curve *C*. 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. The unsatisfactory result of the CV AC model 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. The global minimization of the above energy functional does not provide better result when segment is inhomogenous as shown in **Fig. 2(b)**.

As stated earlier, our methodology is to incorporate an evolving term based on the Bhattacharyya distance to the CV energy functional that not only minimizes the dissimilarities within the object but also maximizes the distance between the two regions.

Bhattacharyya distance is a method of segmentation based on minimizing error probability that employes the concept of distance between probability distributions as the error probability criterion because it can be defined and devised easily. Moreover, it measures the similarity of two discrete or continuous probability distributions. It is the amount of measure of overlie between two statistical samples [22].

The proposed energy function is:

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

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 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).

$\beta \in [0,1]$, $B(C) \equiv \int_x \sqrt{f_1(x)\, f_2(x)}\, dx$ with the Bhattacharyya coefficients [22] are given with Eq. 3 and 4

$$f_1(x) = \frac{K_\sigma(x)^* \left[H(\phi(x))I(x) \right]}{K_\sigma(x)^* \left[H(\phi(x)) \right]} \tag{3}$$

$$f_2(x) = \frac{K_\sigma(x)^* \left[H(\phi(x))I(x) \right]}{K_\sigma(x)^* \left[1 - H(\phi(x)) \right]} \tag{4}$$

where $f_1(x)$ and $f_2(x)$ are the local fitting functions [22], which depend on the level set function ϕ, and need to be updated in each contour evaluation, and $H(\bullet)$ and $\delta'(\bullet) \equiv H'(\bullet)$ respectively the Heaviside and the Dirac functions [22]. 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, $B(C)$ 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

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

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

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

$$\frac{\partial \phi}{\partial t} = \left| \nabla \phi \right| \left\{ \begin{array}{l} \gamma k + \eta + \beta \left[(I - C_{in})^2 - (I - C_{out})^2 \right] - \\ (1-\beta) \left[\begin{array}{l} \frac{B}{2} \left(\frac{B}{A_{in}} - \frac{B}{A_{out}} \right) + \\ \frac{1}{2} \int_\Omega \delta(x-1) \left(\frac{1}{A_{out}} \sqrt{\frac{f_1}{f_2}} - \frac{1}{A_{in}} \sqrt{\frac{f_2}{f_1}} \right) \end{array} \right] \end{array} \right\} \tag{6}$$

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 model's limitation in segmenting inhomogeneous objects as shown in **Fig. 2(c)**.

Thus, the proposed model overcame the limitation of inhomogeneous objects yielding the body detector more robust to illumination changes and noise in general.
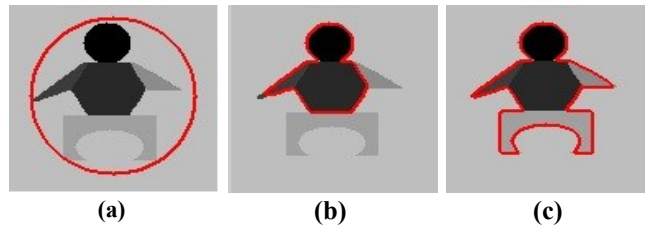
(a)   **(b)**   **(c)**

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

4. Experimental Results

In order to evaluate the performance of the proposed technique, we collected various still activities, such as bending, place jumping, one hand waving, two hand waving, clapping, and boxing, in indoor environment by using depth cameras. During experiments cameras were normal to the human subject, i.e., the angle between the camera and the human was 90°. Each video consisted of several activities, where each activity consisted of several frames and each frame was of the size 144 x 180 pixels.

In the proposed technique, the active contour evolution in a certain frame is performed independently of the other frames, meaning that the human body segmentation in video is done frame-based. The only exploited information is the final contour obtained in the previous frame that is used to determine the initial position of the active contour in the current frame.

First, an ellipse with major axis along *y*-axis of length 25 and minor axis along *x*-axis of length 10 was manually selected closer to the object as the initial contour. Then from the second frame, the position of the initial contour's center in the current frame was the mean value of the points along the final contour in the previous frame. For example, suppose that along the final contour of frame $n (n \geq 1)$, there are M points $\left(x_i^{(n)}, y_i^{(n)}\right)$, $i = 1?M$. Then, the center $\left(c_x^{(n+1)}, c_y^{(n+1)}\right)$ of the initial contour in frame $(n+1)$ is calculated as:

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

As stated earlier, the proposed algorithm not only minimizes the dissimilarities within an object, but also maximizes the distance between the two regions (in our case human body and background), which helps the model to achieve better performance even in the presence of noise. To evaluate this, the salt-pepper image set was created by adding salt and pepper noise to the frames. When applied to this set, the proposed algorithm provided accurate results, which indicates that it is robust to noise in general, for which we also compared the proposed segmentation model with [26] who solved the limitations of CV AC model under the same settings.

The experiments were implemented using Matlab on a PC equipped with 2.5 GHz Dual Core Intel processor and 3 GB RAM. There are three parameters that were used in the proposed approach. First parameter is γ that controls the smoothness of the contour. Smaller γ helps to detect more objects of various sizes, including small points caused by noise. The second parameter is η, which moves the contour along its normal direction. Increasing η will speed up the evolution but may make the contour pass through weak edges. The value of η=0

was used in all experiments for fair comparison. The third parameter is β which weights the constraints of within-object homogeneity and between-object dissimilarity.

4.1 First Experiment

In this experiment, the performance of the proposed AC model for human body segmentation was compared against that of the CV AC model for two video datasets. The results are summarized in **Fig. 3** and **Fig. 4**. Video dataset shown in **Fig. 3** consists of a sequence of three activities, namely place jumping, bending, and two hand waving (three frames per activity), whereas the one in **Fig. 4** consists of clapping and boxing (three frames per activity).

In both of these figures, the first three rows correspond to the segmentation results of the proposed segmentation technique, whereas the last three rows provide to the segmentation results of the CV AC model for the same video sequence. It is obvious from the **Fig. 3** and **4** that the proposed model worked well with activities like place jumping, bending, two hand waving, clapping or boxing. Even with the above mentioned initialization scheme (Section (4.1)), the CV AC model still failed to capture the correct object as shown in the last three rows of **Fig. 3** and **4**.

4.2 Second Experiment

In this experiment, the performance of the proposed AC model was analyzed in the case of noisy data. **Fig.3** and **Fig. 4** show that the performance of the proposed model is much better than that of the CV AC model in normal scenarios because of its better segmentation. As for the cases when noise could be present in the video frames, CV AC model cannot segment the whole body. However, the proposed approach works well in both cases, which means that white noise does not affect its performance. The sample segmentation results for the noisy case are shown in **Fig. 5**, and these results indicate that the proposed segmentation model provides better performance in normal as well as in noisy environments.

4.3 Third Experiment

In this experiment, the proposed AC model has been compared with some state-of-the-art segmentation methdos including [24–27] in order to show the efficacy of the proposed AC model.

The authors of [24] proposed a level set-based method for segmentation; however, in their method, they defined a local and global intensity clustering functions for segmentation based on neighborhood pixels that may fail on depth images because of variation in intensities, such as pixel intensities of the hands in clapping or boxing activities.

Likewise, the authors of [25] proposed a locally statistical AC model based on the means of the Gaussian intensity distributions in the transformed domain by employing a moving window. However, moving window affects the entire curve and consumes extra time to invert the matrix, which can interfere with fast interactive reshaping of a curve [28].

Similarly, in [26, 27], the authors proposed a region-based active contour model (level set) for segmentation and solved the limitations of CV AC model in noisy environment. However, the major limitation of the region-based method is that it cannot provide accurate segmentation results due to the intensity inhomogeneity [24].

In order to solve the limitations of the abovementioned methods, we came up with a new segmentation model that not only minimizes the dissimilarities within the object but also maximizes the distance between the object (human body) and the background. The comparison results are shwon in **Fig. 6**. It can be seen that proposed model performed better than the existing works.
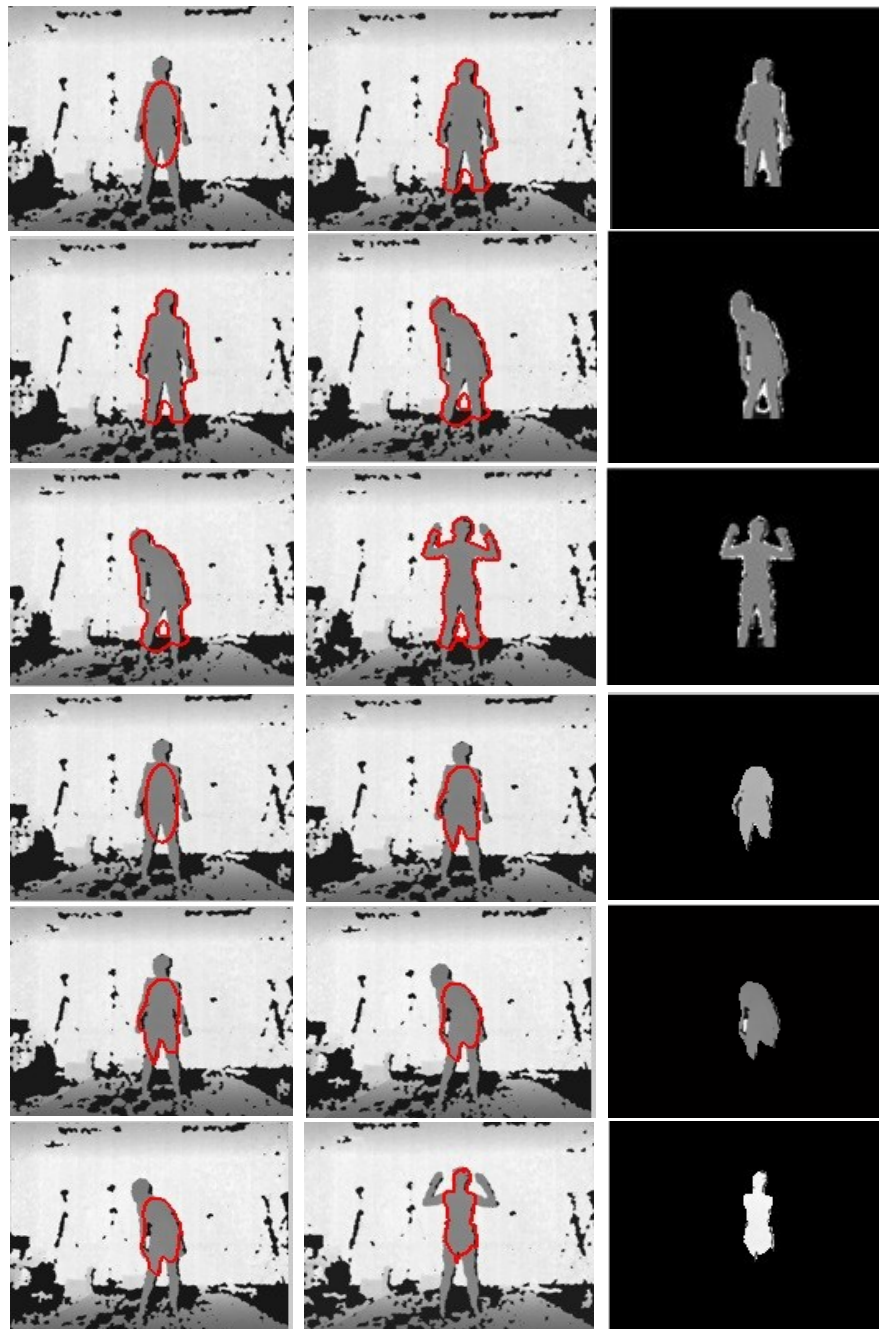
Fig. 3. First three rows show the sample segmentation results of the proposed approach showing accurate human body segmentation on a video sequence of three activities, namely place jumping, bending, and two hand waving, respectively (with $\gamma = 0.5$, $\beta = 0.4$, cpu time = 21.2 s/frame). In each row, the red ellipse in the first image shows the final contour in the previous image whose mass center is used to adjust the contour in the current frame that is shown in the second image, and the last image indicates the segmented human body.

Whereas the remaining three rows (fourth, fifth, and sixth rows) show the sample segmentation results of CV AC model for the same video sequence (with; $\gamma = 0.1$, $\beta = 1.0$, cpu time = 15.31 s/frame). It is obvious that the proposed model worked well but the CV AC model failed.
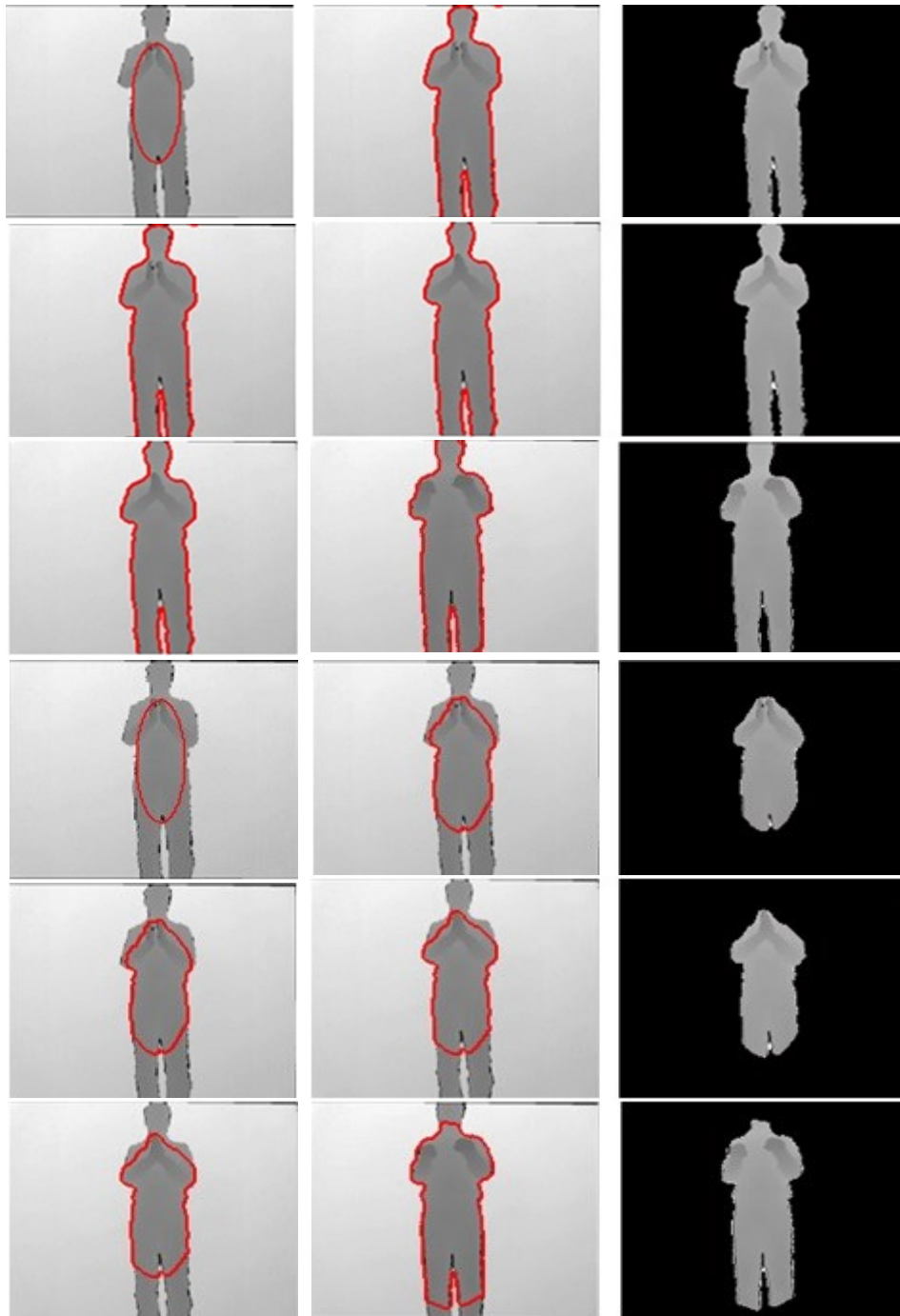
Fig. 4. First three rows show the sample segmentation results of the proposed approach on a video sequence of two activities, namely clapping and boxing (with; γ = 0.4, β = 0.6, cpu time = 30.00 s/frame).
Whereas the remaining three rows (fourth, fifth, and sixth rows) show the sample segmentation results for the CV AC model for the same video sequence (with γ = 1.0, β = 0.6, cpu time = 28.43 s/frame). It is obvious that the CV AC model failed to capture the body correctly as compared to the proposed AC model.
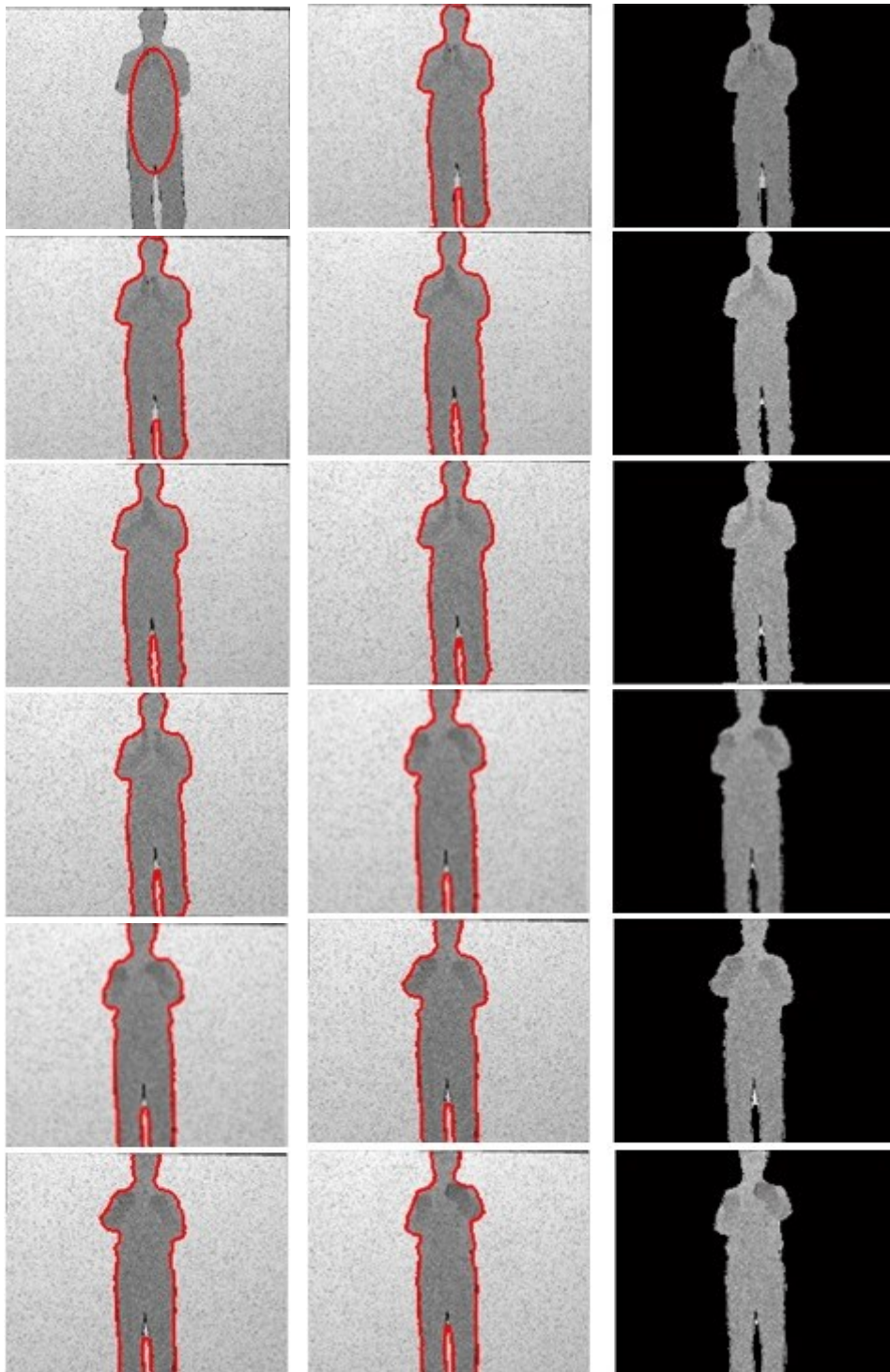
Fig. 5. Sample segmentation results for the proposed AC model on the video sequence of clapping and boxing (with added noise), and with γ = 0.4, β = 0.6, cpu time = 40.23 s/frame. It is obvious that the proposed model worked well in both noisy and normal environments.
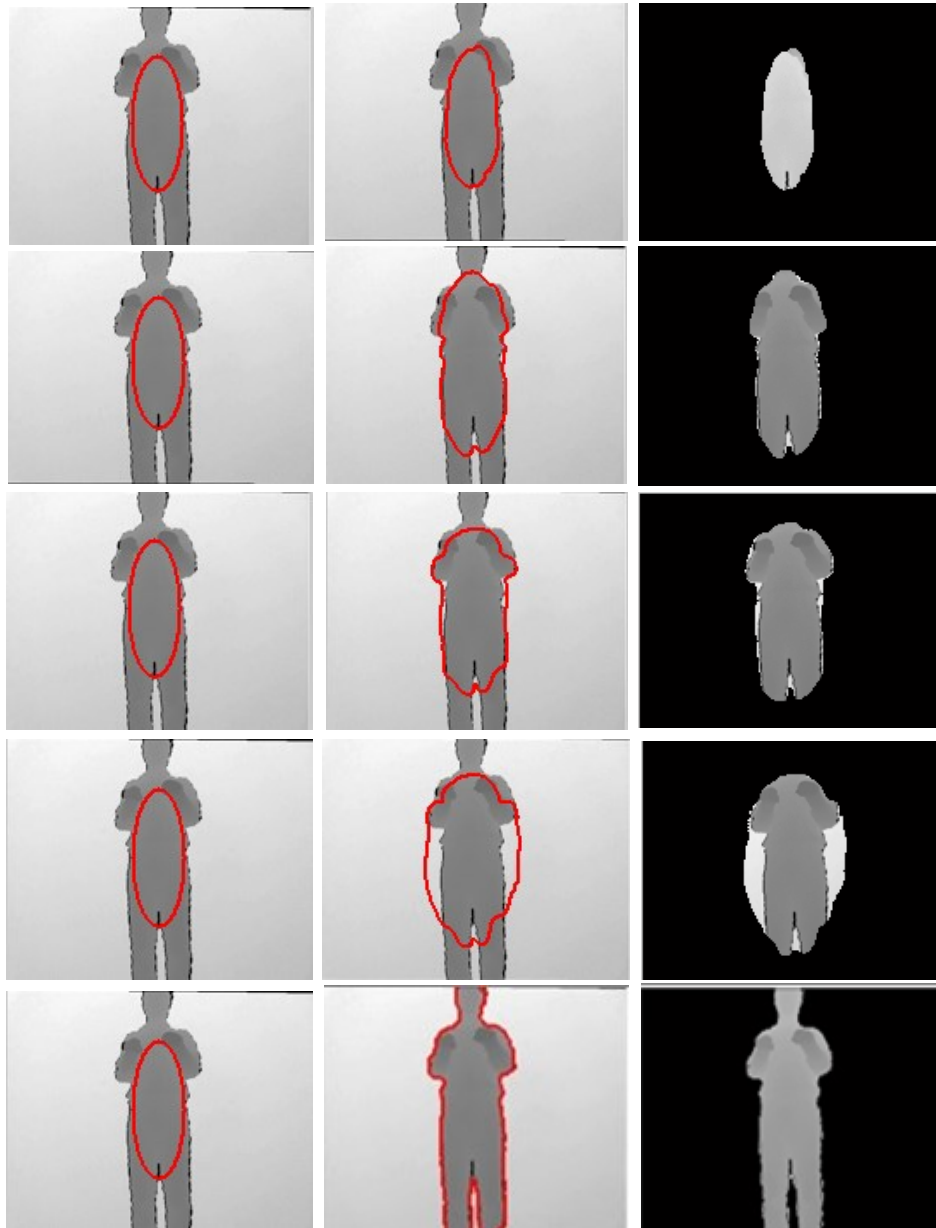
Fig. 6. Comparison results on the video sequence of boxing for the proposed AC model against some state-of-the-art segmentation methods, i.e., [24] (first row), [25] (second row), [26] (third row), [27] (fourth row), and the last row provides the sample segmentation results of the proposed AC model. It is obvious that the proposed model works well against the existing state-of-the-art methods.

In this paper, an active contour model is presented for human body segmentation from depth images. Compared to the conventional CV AC model and other state-of-the-art methods for static image segmentation, the proposed model is not only more accurate but it is also more robust to noise.

Like other AC models, when applied to depth data where the background environment is much more arbitrary, it requires a less relaxed initialization scheme, i.e., the initial contour

should be manually selected close to the object in order to correctly converge. However, from that point onwards, unlike CV AC model, the proposed model uses an efficient and straightforward way to locate the contour's position in the next frame, i.e., 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. This way works well with static activity video where the displacement of object between consecutive frames is small.

Another reason for better performance of the proposed model is the fact that it is based on the combination of the CV energy function and the Bhattacharya distance between the density functions inside and outside the curve. The subsequent flow examines for a segmentation that not only minimizes the dissimilarities within the object but also maximizes the distance between the object and the background. The evolution flow of the proposed model is derived using level-set framework, making it inherit advantages of a geometric active contour model.

As shown in **Fig. 5**, the proposed segmentation technique is less sensitive to noise. This is due to the fact the model employs a different value for the parameter β and γ in the case of noisy data. In order to investigate the sensitivity with respect to β and γ, we applied the proposed AC model on the noisy image data using different values for β and γ, and found $\beta = 0.6$ and $\gamma = 0.4$ to be the most optimal values for the case of noisy data. Employing different values of β and γ for the noisy case works well for the proposed AC model, but not for the CV AC model, because a higher γ makes the CV AC model incapable to capture the fine-scale structure of the object.

It is also possible to see (see Fig. captions) that the CPU time of the proposed AC model is comparable to that of the CV AC model despite the higher computational cost. This is due to the fact that the extra evolving term helps to move the curve faster towards convergence. As a conclusion, the proposed AC model is a good candidate for unsupervised image segmentation where the images are commonly noisy.

The proposed active contour model is a feasible choice for many applications where the number of subjects is limited, such as patient monitoring. However, if there are a large number of subjects present in a video frame, individual initial contour will be required for each subject. Furthermore, the developed model fails to segment the human body for the dynamic activities such as running, walking, jogging. Because, most of these activities consist of motion information for which the final contour of the previous frame cannot be moved to the next frame in the proposed model. Also, there is a large displacement of object between consecutive frames of those videos, making the previous-frame-based initial position in the current frame far from the object of interest. Therefore, further research is needed to modify the proposed active contour model with the integration of the motion information to segment the still as well as dynamic activities from depth.

Moreover, in the proposed model, we first initialize the initial contour manually that should be near to the human body. The proposed model might not work well if the contour is far away from the human body. Therefore, further research is also needed to make the proposed algorithm automated. Nevertheless, its superior performance against existing methods indicates the feasibility of using the proposed segmentation scheme for still body human segmentation from depth images.

5. Conclusion

The accuracy of video based human activity recognition extensively depends on the performance of employed human body segmentation technique. In this paper, an active contour model with the integration of the CV energy function and the Bhattacharya distance

function is proposed that not only minimizes the dissimilarities within the object but also maximizes the distance between the object and the background. The proposed model is more robust to noise as compared to conventional CV AC model. However, like other active contour models, when applied to depth data where the background atmosphere is much more random, it requires a less relaxed initialization scheme, such as the initial contour should be close to the flow of the current frame and the mass center of those new objects in order to correctly converge. Then this mass center of points along the final contour in the previous frame is used as the center of the initial contour in the current frame. By this way the human body segments very accurately in static activities where the displacement of object between consecutive frames is small.

References

[1] C. Hoffman and D. Rice, "Chronic care in america: A 21st century challenge," *Princeton, NJ: The Robert Wood Johnson Foundation*, 1996.

[2] E. M. Tapia, S. S. Intille, and K. Larson, "Activity Recognition in the Home Using Simple and Ubiquitous Sensors," *In Pervasive*, pp. 158 – 175, 2004.

[3] N. P. Tiilikainen, "A comparative study of active contour snakes," *Copenhagen University, Denmark*, 2007.

[4] K. S. Ntalianis, N. D. Doulamis, A. D. Doulamis, and S. D. Kollias, "An active contour-based video object segmentation scheme for stereoscopic video sequences," in *Proc. of Electrotechnical Conference, 2000. MELECON 2000. 10th Mediterranean*, vol. 2, pp. 554–557, IEEE, 2000.

[5] M. Kunt, A. Ikonomopoulos, and M. Kocher, "Second-generation image-coding techniques," *Proceedings of the IEEE*, vol. 73, no. 4, pp. 549–574, 1985.

[6] F. Meyer and S. Beucher, "Morphological segmentation," *Journal of visual communication and image representation*, vol. 1, no. 1, pp. 21–46, 1990.

[7] D.W. Murray and B. F. Buxton, "Scene segmentation from visual motion using global optimization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 2, pp. 220–228, 1987.

[8] A. D. Doulamis, N. D. Doulamis, K. S. Ntalianis, and S. D. Kollias, "Unsupervised semantic object segmentation of stereoscopic video sequences," in *Proc. of Information Intelligence and Systems, 1999. Proceedings. 1999 International Conference on*, pp. 527–533, IEEE, 1999.

[9] C.-Y. Huang and M.-J. Wu, Image Segmentation. *ECE 533 Final Project. University of Wisconsin- Madison*, 2006.

[10] F. Meyer, "Color image segmentation," in *Proc. of Image Processing and its Applications, 1992., International Conference on*, pp. 303–306, IET, 1992.

[11] T. F. Chan and L. A. Vese, "Active contours without edges," *Image Processing, IEEE Transactions on*, vol. 10, no. 2, pp. 266–277, 2001.

[12] M. H. Coen et al., "Design principles for intelligent environments," in *Proc. of the National Conference on Artificial Intelligence*, pp. 547–554, JOHN WILEY & SONS LTD, 1998.

[13] C. D. Kidd, R. Orr, G. D. Abowd, C. G. Atkeson, I. A. Essa, B. MacIntyre, E. Mynatt, T. E. Starner, and W. Newstetter, "The aware home: A living laboratory for ubiquitous computing research," in *Proc. of Cooperative buildings. Integrating information, organizations, and architecture*, pp. 191–198, Springer, 1999.

[14] J. K. Aggarwal and Q. Cai, "Human motion analysis: A review," *Computer Vision and Image Understanding*, vol. 73, no. 3, pp. 428–440, 1999.

[15] M. Siddiqi, M. Fahim, S. Lee, and Y.-K. Lee, "Human activity recognition based on morphological dilation followed by watershed transformation method," in *Proc. of Electronics and Information Engineering (ICEIE), 2010 International Conference On*, vol. 2, pp. V2–433, IEEE, 2010

[16] M. Z. Uddin, J. Lee, and T.-S. Kim, "Shape-based human activity recognition using independent component analysis and hidden Markov model," *in Proc. of New Frontiers in Applied Artificial Intelligence*, pp. 245–254, Springer, 2008.

[17] M. Z. Uddin, J. Lee, and T.-S. Kim, "Independent shape component-based human activity

recognition via hidden markov model," *Applied Intelligence*, vol. 33, no. 2, pp. 193–206, 2010.

[18] M. Z. Uddin and T.-S. Kim, "Continuous hidden markov models for depth map-based human activity recognition," *Hidden Markov Models, Theory and Applications*, pp. 225–247, 2011.

[19] M. Z. Uddin, T.-S. Kim, and J. T. Kim, "Video-based indoor human gait recognition using depth imaging and hidden markov model: a smart system for smart home," *Indoor and Built Environment*, vol. 20, no. 1, pp. 120–128, 2011.

[20] A. Jalal, M. Z. Uddin, J. T. Kim, and T.-S. Kim, "Recognition of human home activities via depth silhouettes and transformation for smart homes," *Indoor and Built Environment*, vol. 21, no. 1, pp. 184–190, 2012.

[21] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988.

[22] T. Kailath, "The divergence and bhattacharyya distance measures in signal selection," *Communication Technology, IEEE Transactions on*, vol. 15, no. 1, pp. 52–60, 1967.

[23] R. Rusyaizila, Z. Nasriah, and S. Putra, "Privacy issues in pervasive healthcare monitoring system: a review," *in World Academy of Science, Engineering and Technology*, pp. 741–747, 2010.

[24] C. Li, R. Huang, Z. Ding, C. Gatenby, D. Metaxas, and J. C. Gore, "A Level Set Method for Image Segmentation in the Presence of Intensity Inhomogeneities with Application to MRI," *Image Processing, IEEE Transactions on*, vol. 20, no. 7, pp. 2007–2016, 2011.

[25] K. Zhang, L. Zhang, K. M. Lam, and D. Zhang, "A Locally Statistical Active Contour Model for Image Segmentation with Intensity Inhomogeneity," *arXiv preprint arXiv:1305.7053*, 2013.

[26] M. F. Talu, "ORACM: Online region-based active contour model," *Expert Systems with Applications*, vol. 40, no.6, pp. 6233–6240, 2013.

[27] K. Zhang, L. Zhang, H. Song and W. Zhou, "Active contours with selective local or global segmentation: A new formulation and level set method," *Image Vis. Comput.,* vol. 28, no. 4, pp. 668–676, 2010.

[28] V. Hadziavdic. "A Comparative Study of Active Contour Models for Boundary Detection in Brain Images". Diploma Project. Faculty for Mathematical and Natural Sciences, University of Tromso, 1999.