TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.5, May 2014, pp. 3521 ~ 3528
DOI: http://dx.doi.org/10.11591/telkomnika.v12i5.3539

A Coarse-to-Fine Human Body Segmentation Method in Video

Yingna Deng*, Wenqing Wang
Xi'an University of Posts and Telecommunications,
Weiguo Road, Xi'an, China, 86-029-88166461
*Corresponding author, e-mail:dengyingna@126.com

Abstract

Human body precise segmentation is difficult because of inter-occlusion when there are multiple human bodies in video. A coarse-to-fine segmentation method is proposed. In coarse segmentation, human shape models are used to get human's position and coarse region. The human models with variant scale and posture are constructed with head, torso, and legs. For each human body, its corresponding human shape model is obtained by model matching, and by which human position is obtained roughly. Human precise contour is obtained in fine segmentation by curve evolution with initial contour obtained from coarse segmentation. Experiment results show that the proposed method could segment human object precisely.

Keywords*: Human shape model, level set, bayesian model, segmentation*

1. Introduction

Crowd precise segmentation is important for human tracking and recognition. However, when there are multiple pedestrians, human object precise segmentation is difficult because of inter-occlusion.The models which human body segmentation based on could be classified into dynamic model and appearance model. Dynamic models are made up of object's moving speed, direction and other dynamic features [1, 2]. Appearance models are made up of color, shape, position and other features which represent object's appearance. Elgammald and Ramanan used color and position information to construct object models under the assumption of object entering the camera view lonely [3, 4]. and in this condition, each object's region was obtained by kernel density estimation. Zhe modified Elgammald's method, and used EM evolution for human segmentation [5]. Wu and Sapp made up human body part classifications with boosting algorithm [6, 7]. Lu employed a coarse to fine method to segment human body from photos [8].

Human shape models coud provide position, area, posture, and some other rough information; it means that based on human shape models, human object could be segmented roughly. However, getting precise contour of each object is necessary for human object precise segmentation. In this paper, a method based on both human shape model and level set is proposed. Human objects' number, rough area and shape models are obtained through initial segmentation, then on the basis of initial segmentation, each object's precise contour is obtained by curve evolution using level set.

2. Coarse Segmentation Based on Human Shape Model

2.1. Overview of the Method

Human body is made up of head, tarso, arms and legs, and the human shape changes regularly, as shown in Figure 1 and Figure 2. So, 7 human shape models with different posture are consturcted consisting of ellipses to simulate human walking including 3 front views and 4 side views, as shown in Figure 3.

Suppose the object region detected by background subtraction is denoted as I, θ is human shape model, the best segmentation result could be obtained by estimating the maximum of posterior probability, as shown in Equation (1).

Received June 18, 2013; Revised December 11, 2013; Accepted December 31, 2013

$$\theta^{*} = \arg\max_{\theta\in\Theta} P(\theta \mid I) \tag{1}$$

Where $\theta=\{n,\{M_1,M_2,\ldots,M_n\}\}$, n is the number of human objects, M is the human model parameter, defined as $M=\{h,height,l\}$, and each factor presents head position, human height and human pose separately.

Based on Bayesian theory, we know that,

$$P(\theta \mid I) \propto P(I \mid \theta)P(\theta) \tag{2}$$

Where $P(\theta)$ is the object prior probability, $P(I|\theta)$ is the similarity of model and foreground region. Suppose each object appearances with equal prior probability, the segmentation result is θ^* which maximize the similarity $P(I|\theta)$.

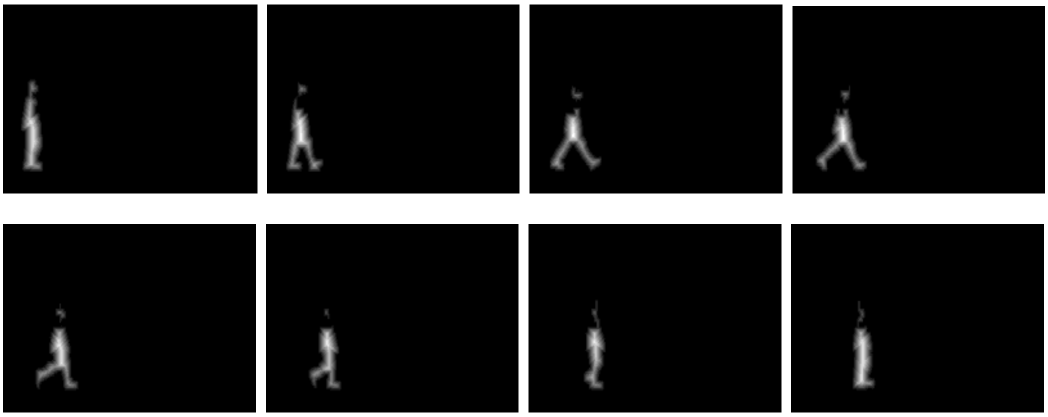


Figure 1. Side View of Human Walking

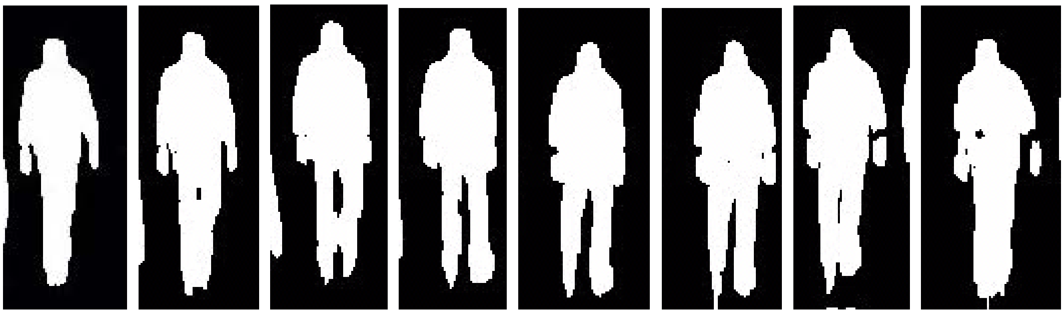


Figure 2. Front View of Human Walking

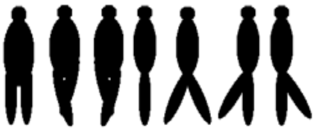


Figure 3. Human Pose Models

2.2 Head Candidates

2.2.1. Get Head Candidate Region with Region Search

For a point h in foreground, let it as the upper left corner to build a rectangle with height of object and width of 1/3 height. Section 2.3 describes how to get the height of human shape model. The probability of head candidate point *r*(*h*) is the proportion of pixels in foreground and the rectangle, that is:

$$r(h)=\frac{S_{obj}}{S_{rec}} \tag{3}$$

Based on Equation (3), object head candidates could be obtained by setting a threshold *T*, as shown in Figure 4(c). However, the head candidates with high probabilities are not all true head candidates, so object edge and head shoulder matching is applied to eliminate most of the false candidates [9].

2.2.2. Edge Based Head Shoulder Matching

The head shoulder shape is almost unchanged while human is walking, so a head shoulder model is constructed to evaluate the head candidates, as shown in Figure 4. For a point h in edge image, the similarity between it and the head shoulder model is defined as follows:

$$P(h)=\sum_{v} G(v,v_0,\delta)n_0(v)gn_m(v)\Big/N \tag{4}$$

Where v is the nearest point of edge image to the head shoulder model in the direction that the test line points to, v_0 is the intersecrion of testline and head shoulder model contour, $G(v,v_0,\delta)$ is a Gaussian model, n_0 (v) is the normal direction of point v, and $n_m(v)$ is the direction of test line where point v is, N is the number of test lines in head shoulder model.
The method of get head candidates is shown in Figure 5.

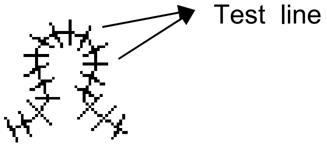


Figure 4. Head Shoulder Model

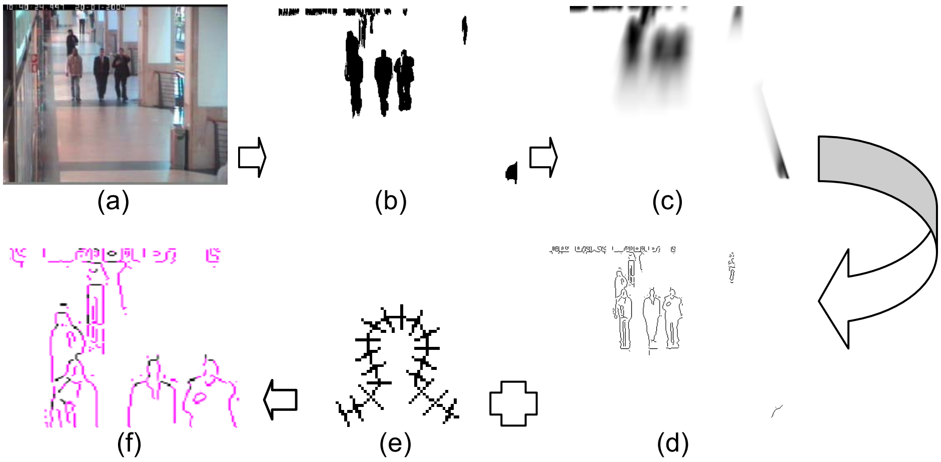


Figure 5. The Process of Getting Head Candidates. (a) a video frame, (b) object region, (c) probability of head candidate, (d) object edge image, (e) head shoulder model, (f) head candidates (black points).

2.3. Human Shape Model Height Estimation

2.3.1. Foot Point Estimation

The height of object in image is related to the distance between object and camera optical center, so, the human model height could be estimated from the head point *h* by estimating the foot point under the assumption that all the humans are all adult, that is:

$$\begin{bmatrix} x_f \\ y_f \\ 1 \end{bmatrix} = Hg \begin{bmatrix} x_h \\ y_h \\ 1 \end{bmatrix} \tag{5}$$

Where *H* is the homography matrix between the planes of feet and heads are located, and it could be estimated with least squares method [10]. (x_f,y_f) is the estimated foot point coordinate, and (x_h,y_h) is the head point coordinate.

2.3.2. Homograpy Matrix Estimation

Homography matrix connect human head from foot points. Given *n*(>=4) head points and their corresponding foot points (m_1^i, m_2^i),the homography matrix *H* could be estimated.

Let $H = \begin{bmatrix} h_1, h_2, h_3 \\ h_4, h_5, h_6 \\ h_7, h_8, 1 \end{bmatrix}$, and it could be written as a vector $h = (h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, 1)$, given a pair of matching points $m_1 = [x_1, y_1, 1]^T$ and $m_2 = [x_2, y_2, 1]^T$, 2 linear equations about h could be obtained, as shown in Equation (6) and Equation (7).

$$(x_1, y_1, 1, 0, 0, 0, -x_2 x_1, -x_2 y_1)h = x_2 \tag{6}$$

$$(0, 0, 0, x_1, y_1, 1, -y_2 x_1, -y_2 y_1)h = y_2 \tag{7}$$

There are 8 unknown numbers in *H*, so 4 pairs of matching points are needed to get *H*, besides, the matching points should not be colinear.

2.4. Similarity between Object Region and Shape Model

When shape model covers object region, the similarity between image *I* and shape model θ is defined as Equation (8).

$$P(I \mid \theta) = \alpha e^{-(\lambda_{10} N_{10} + \lambda_{01} N_{01})} \tag{8}$$

where α is a constant independent of θ. N_{10} is the number of pixels that are in object region but not in shape model, N_{01} is the number of pixels that are in shape model but not in object region. λ_{10} is a coefficient dependent on the probability that a pixel is in object region but not in shape model, and λ_{01} is a coefficient dependent on the probability that a pixel is not in object region but in shape model.

2.5. Human Shape Model Matching

Suppose there are *N*s head candidates obtained from section 2.2, the true object number *n* meet the condition of *n*≤*N*s. The solution *θ** which maximized the probability *P*(*I*|*θ*) is obtained by iteration.

For a head candidate point h, the pose l in object model is obtained by Equation (9).

$$l = \arg\max_{j=1:7} P(l_j), P(l_j) = \frac{S_{Fj}}{S_{Mj}} \tag{9}$$

Where S_M is model area, S_F is foreground area covered by the model.

The steps of finding solution θ* which maximized posteriors are as follows:

(1) Initialization，let $\theta^*=\{Ns,\{M_1,M_2,\ldots,M_{Ns}\}\}$,$M_i=\{hi,height_i,l_i\}$, the similarity between model and object region $P(I|\theta^*)$ could be computed through Equation (8).

(2) For all the Ns head candidates, put off the current object t（$t=1,2,\ldots,N_s$），let $\theta_{cur}=\theta^*-\theta_t$, $\theta_t=\{1,M_t\}$, compute the similarity $P(I|\theta_{cur})$ again.

(3) If $P(I|\theta_{cur})>P(I|\theta^*)$, let $\theta^*=\theta_{cur}$, else，keep θ^* unchanged, and return to step (3).

3. Fine Segmentation Based on Level Set

3.1. Overview of the Method

The contour is initialized based on the coarse area of each object obtained from initial segmentation, and the precise region is obtained through curve evolution with level set, that is let N-1 level sets to achieve N regions segmentation.

Suppose the image is separated into N regions by N-1 curves without region overlap, let the ith evolution curve is expressed as $\vec{\gamma}_i(s,t), i\in[1,\ldots N-1]$,where arc length $s\in[0,1]$, t is evolution time, the close curve set expressed as $\left\{\vec{\gamma}_i(s,t):[0,1]\rightarrow\Omega\right\}_{i=1}^{N-1}$ could separate the image Ω into R regions, as shown in Equation (10).

$$R_1=R_{\vec{\gamma}_1},R_2=R^c_{\vec{\gamma}_1}\cap R_{\vec{\gamma}_2},\ldots,R_{N-1}=R^c_{\vec{\gamma}_1}\cap R^c_{\vec{\gamma}_2}\cap\ldots\cap R_{\vec{\gamma}_{N-1}},\ldots,R_N=R^c_{\vec{\gamma}_1}\cap R^c_{\vec{\gamma}_2}\cap\ldots\cap R^c_{\vec{\gamma}_{N-1}} \tag{10}$$

Where $R_{\vec{\gamma}_i}$ and $R^c_{\vec{\gamma}_i}$ indicate inside and outside region separately, zero level sets number N-1 is human objects number obtained from initial segmentation in this paper.

3.2. Realization of Human Body Segmentation with Level Set

Suppose the image is $I:\Omega\subset R^2$, the energy is defined as:

$$E(\vec{\gamma},u)=E^R(\vec{\gamma},u)+E^E(\vec{\gamma})+E^C(\vec{\gamma}_i) \tag{11}$$

Where:

$$E_R(\vec{\gamma}_i,u)=\lambda_1\int_{R_{\gamma_1}}(I(x)-u_{R_{\gamma_1}})^2dx+\lambda_2\int_{R^c_{\gamma_1}\cap R_{\gamma_2}}(I(x)-u_{R^c_{\gamma_1}\cap R_{\gamma_2}})^2dx+\ldots+\lambda_n\int_{R^c_{\gamma_1}\cap R^c_{\gamma_2}\cap\ldots\cap R^c_{\gamma_{N-1}}}(I(x)-u_{R^c_{\gamma_1}\cap R^c_{\gamma_2}\cap\ldots\cap R^c_{\gamma_{N-1}}})^2dx \tag{12}$$

$$E^E(\vec{\gamma})=\mu\sum_{i=1}^{N-1}\int_{\vec{\gamma}_i}ds \tag{13}$$

$$E^C(\vec{\gamma}_i)=\frac{v}{2}\sum_{i=1}^{N-1}\int_{\Omega}(|\nabla\vec{\gamma}_i|-1)^2dx \tag{14}$$

Where $\vec{\gamma}=\left\{\vec{\gamma}_i,i=1,\ldots,N-1\right\},u=\{u_{Ri},i=1,\ldots,N\}$ is the pixel average value, $\lambda_i,i=1,2,\ldots,N$ is weight value, and μ is weight value above 0.

So, the human object precise segmentation is to find the minimum energy $E(\vec{\gamma},u)$.

Suppose the zero level set according to the region encircled by evolution curve sets $\left\{\vec{\gamma}_i,i=1,\ldots,N-1\right\}$ is $\{\phi_i(x,y)=0,i=1,2,\ldots,N-1\}$, and the level set function is defined as follows:

$$\begin{cases}\phi_i(x,y)>0 & x\in Inside(\vec{\gamma}_i)\\ \phi_i(x,y)=0 & x\in(\vec{\gamma}_i)\\ \phi_i(x,y)<0 & x\in Outside(\vec{\gamma}_i)\end{cases} \tag{15}$$

Suppose H(x) is Heaviside function, and it is defined as follows:

$$H(x)=\begin{cases}1, if\ \ x\geq 0\\ 0, if\ \ x<0\end{cases} \tag{16}$$

The region indicative functions are defined as follows:

$$\begin{cases}\chi_{R_1}=\chi_{R\vec{\gamma}_1}=H(\phi_1)\\ \chi_{R_2}=\chi_{R\vec{\gamma}_1^c}\chi_{R\vec{\gamma}_2}=[1-H(\phi_1)]H(\phi_2)\\ \cdots\\ \chi_{R_N}=\chi_{R\vec{\gamma}_1^c}\chi_{R\vec{\gamma}_2^c}\cdots\chi_{R\vec{\gamma}_{N-1}^c}=\prod_{i=1}^{N-1}[1-H(\phi_i)]\end{cases} \tag{17}$$

So, the curve evolution equation could be expressed by Equation (18).

$$\begin{cases}\dfrac{\partial\phi_1}{\partial t}=-\left[(I(x)-u_{R_1})^2-\Phi_1+\mu k_1\right]\left\|\overrightarrow{\nabla\gamma_1}(x,t)\right\|+v(\nabla\vec{\gamma}_1^{\,2}-k_1)\\ \dfrac{\partial\phi_2}{\partial t}=-\left[(I(x)-u_{R_1})^2-\Phi_2+\mu k_2\right]\left\|\overrightarrow{\nabla\gamma_2}(x,t)\right\|+v(\nabla\vec{\gamma}_2^{\,2}-k_2)\\ \ldots\\ \dfrac{\partial\phi_{N-1}}{\partial t}=-\left[(I(x)-u_{R_{N-1}})^2-\Phi_{N-1}+\mu k_{N-1}\right]\left\|\overrightarrow{\nabla\gamma_{N-1}}(x,t)\right\|+v(\nabla\vec{\gamma}_{N-1}^{\,2}-k_{N-1})\end{cases} \tag{18}$$

Where u_{R_i} is the average value of region encircled by curve γ_i, and $\Phi_i(x)$ is defined as:

$$\begin{aligned}\Phi_i(x)=&(I(x)-u_{R_{i+1}})^2\chi_{\phi_{i+1}(x,t)>0}(x)+(I(x)-u_{R_{i+2}})^2\chi_{\phi_{i+1}(x,t)<0}(x)\chi_{\phi_{i+1}(x,t)>0}(x)+\ldots\\ &+(I(x)-u_{R_{N-1}})^2\chi_{\phi_{i+1}(x,t)<0}(x)\chi_{\phi_{i+2}(x,t)<0}(x)\ldots\chi_{\phi_{N-2}(x,t)<0}(x)\chi_{\phi_{N-1}(x,t)>0}(x)\\ &+(I(x)-u_{R_N})^2\chi_{\phi_{i+1}(x,t)<0}(x)\chi_{\phi_{i+2}(x,t)<0}(x)\ldots\chi_{\phi_{N-2}(x,t)<0}(x)\chi_{\phi_{N-1}(x,t)<0}(x)\end{aligned} \tag{19}$$

Where $\chi_{\phi_i(x,t)}$ satisfies：

$$\begin{cases}\chi_{\phi_i(x,t)>0}=H(\phi_i(x,t)) & \text{if}\quad \phi_i(x,t)>0\\ \chi_{\phi_i(x,t)<0}=1-H(\phi_i(x,t)) & \text{if}\quad \phi_i(x,t)<0\end{cases} \tag{20}$$

3.3. Crowd Segmentation Step with Level Set

The crowd segmentation steps are as follows:

(1) Level set initialization

Suppose there are *N* human shape models, for each object, the initial curve is a circle which let the center of object model as center point, and let one tenth of human shape model as radius. So, the ith level set is defined as Equation (21):

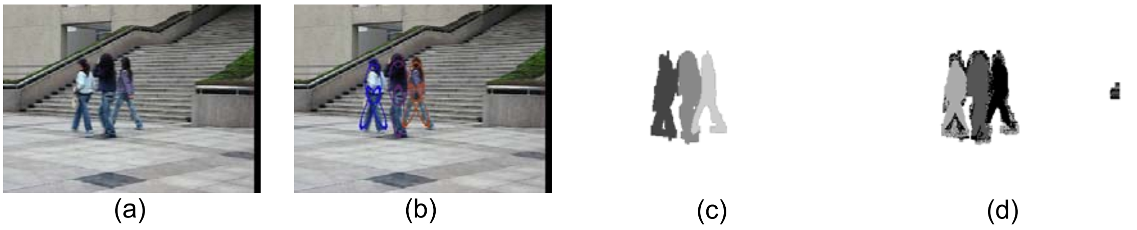
(2) Update the level sets based on Equation (18) sequentially.

(3) If the evolution is unfinished, return to step 2.

$$\begin{cases}\phi_i(x,0)>0 & if\ \ d(x,center_i)<radius_i\\ \phi_i(x,0)=0 & if\ \ d(x,center_i)=radius_i\\ \phi_i(x,0)<0 & if\ \ d(x,center_i)>radius_i\end{cases} \tag{21}$$

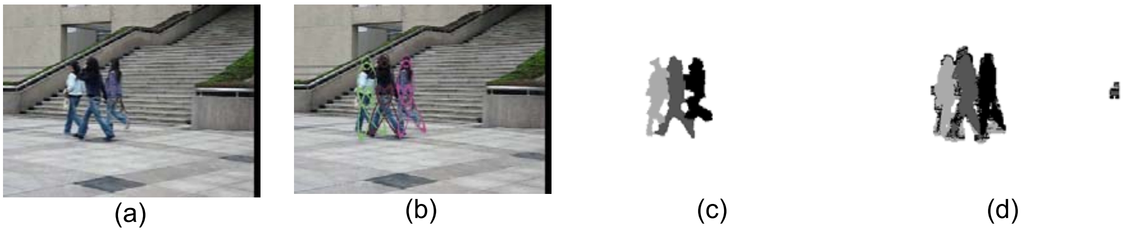
4. Results and Analysis

In order to test the effectiveness of proposed method, videos of CAVIAR and Campus are tested separately. Campus is a video shoot by the authors themselves, as shown in Figure 6-7. CAVIAR is an open video database, and the experimental results as shown in Figure 8-9.



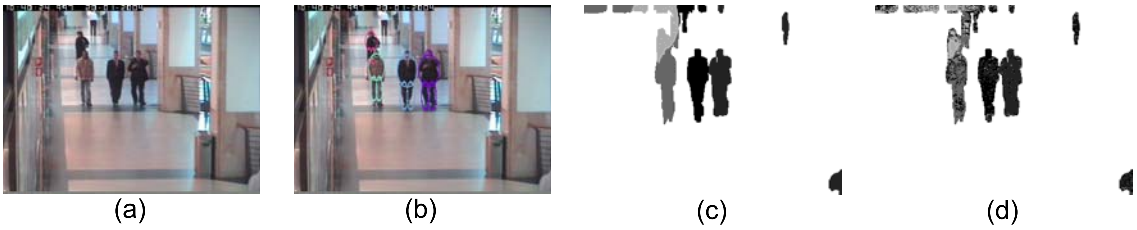
(a) (b) (c) (d)

Figure 6. Object segmentation result of Campus1 frame 720#. (a) shows the original video frame. (b) shows the object models obtained. (c) shows the object region segmented by level set. (d) shows the segmentation result of reference [3]



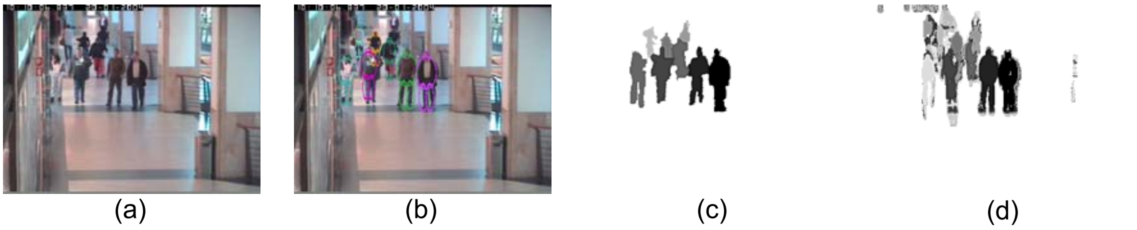
(a) (b) (c) (d)

Figure 7. Object Segmentation Result of Campus1 Frame 725#. (a) shows the original video frame. (b) shows the object models obtained. (c) shows the object region segmented by level set. (d) shows the segmentation result of reference [3]



(a) (b) (c) (d)

Figure 8. Object Segmentation Result of CAVIAR ShopAssistant2Cor Frame 231#. (a) shows the original video frame. (b) shows the object models. (c) shows the object region segmented by level set (d) shows the segmentation result of reference [3]



(a) (b) (c) (d)

Figure 9. Object Segmentation Result of CAVIAR ShopAssistant3Cor Frame 171#. (a) shows the original video frame, (b) shows the object models, (c) shows the object region segmented by level set, (d) shows the segmentation result of reference [3]

In order to test the accuracy of proposed method, we define the accuracy rate as follows:

$$r = \frac{N_{ture}}{N_{obj}} \qquad (22)$$

Where N_{ture} is the pixel number of being segmented correctly for all objects, and N_{obj} is the pixel number of all objects. Table 1 shows the accuracy rate.

Table 1. Human Object Segmentation Accuracy Rate

Method	Accuracy rate
Proposed method in this paper	85.6%
Method in Reference[3]	74.2%

The proposed method could segment crowd object precisely, however, when the color of human varies frequently, the method is not suitable as well. So, a method tolerate to color variation should be researched in the future.

5. Conclusion

When there are multiple human objects, it is difficult to segment each body precisely. A coarse-to-fine segmentation method is proposed in this paper. In coarse segmentation step, a Bayesian estimation based object initial segmentation was done by shape model matching, from which, human's position, height, and posture are obtained roughly. In fine segmentation step, the precise region of each object was obtained through curve evolution with level set. Experimental results show that the proposed method could segment crowd object precisely. For curve evolution of each object, it is expressed by only one level set, so when there are some different colors in object cloth, the segmentation result is not satisfied, so, multiple level sets for one object could be considered in the future.

Acknowledgements

Project supported by the ShaanXi Province Education Department Project (11JK0990) and National Science Foundation of China (61100165)

References

[1] R Yang, Q Zheng. *Multi-moving people detection from binocular sequences.* International Conference on Acoustic, Speech and Signal Processing. Hong Kong.2003 Vol.2:297-300.

[2] JG Brostow, R Cipolla. *Unsupervised Bayesian Detection of Independent Motion in Crowds.* IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York. 2006; 1: 594-601.

[3] AM Elgammal, LS Davis. *Probabilistic framework for segmenting people under Occlusion.* The 8th IEEE Conference on Computer Vision.Vancouver. 2001; 2:145-152.

[4] D Ramanan, DA Forsyth. *Finding and tracking people from the bottom up.* IEEE Conference on Computer Vision and Pattern Recognition. Madison. 2003; 2: 467-474.

[5] Z Lin, SD Larry, D Doermann, et.al. *An interactive approach to pose-assisted and appearance-based segmentation of humans.* Workshop on Interactive Computer Vision. Rio de Janeiro. Brazil. 2007,1-8

[6] Wu Bo, Ram Nevatia. Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors. *International Journal of Computer Vision.* 2007; 27(2): 247-266

[7] B Sapp, A Toshev, B Taskar. *Cascaded models for articulated pose estimation*. European Conference on Computer Vision. Heraklion. 2010; 2: 406.

[8] H Lu, G Fang, X Shao, X Li. Segmenting human from photo images based on a coarse–to-fine scheme. IEEE Transactions on systems, man, and cybernetions–Part B: Cyberentics. 2012; 42(3): 889-899.

[9] Zhao Tao, Ram Nevatia. Bayesian human segmentation in crowded situations. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.Madison. 2003; 2: 459-466.

[10] Zhe LS Davis, D Doermann, D Dementhon. *Hierarchical part-template matching for human detection and segmentation.* The 11th International Conference on Computer Vision. Rio de Janeiro. 2007; 1-8.

[11] AR Mansouri, A Mitiche, C Vazquez. Multiregion competition: A level set extension of region competition to multiple region image partitioning. *Computer Vision and Image Understanding*. 2006; 101(3):137-150.