

Adaptive Silhouette Extraction In Dynamic Environments Using Fuzzy Logic

Xi Chen, Zhihai He, James M. Keller, Derek Anderson, and Marjorie Skubic

Department of Electrical and Computer Engineering
University of Missouri, Columbia, MO 65203, USA

Abstract - Extracting a human silhouette from an image is the enabling step for many high-level vision processing tasks, such as human tracking and activity analysis. Although there are a number of silhouette extraction algorithms proposed in the literature, most approaches work efficiently only in constrained environments where the background is relatively simple and static. In a previous paper, we addressed some of the challenges in silhouette extraction and human tracking in a real-world unconstrained environment where the background is complex and dynamic. We extracted features from image regions, accumulated the feature information over time, fused high-level knowledge with low-level features, and built a time-varying background model. A problem with our system is that by adapting the background model, objects moved by a human are difficult to handle. In order to reinsert them into the background, we run the risk of cutting off part of the human silhouette, such as in a quick arm movement. In this paper, we develop a fuzzy logic inference system to detach the silhouette of a moving object from the human body. Our experimental results demonstrate that the fuzzy inference system is very efficient and robust.

1. INTRODUCTION

Silhouette extraction, namely, segmenting the human body from the background, is often the first step and also the enabling step for many high-level vision analysis tasks, such as human tracking and activity recognition. One central task in silhouette extraction is background modeling. Once a background model is established, those image regions with significantly different characteristics from the background are considered foreground objects. In [1], an LMeds (Least Median of Squares) method is used to construct the background model. The differencing function proposed in [2] is used to extract moving human silhouettes. In [3], a non-parametric kernel density estimation method is employed to model and subtract the background. The adaptive background updating in that approach is able to deal with small motion in the background scene, such as moving tree branches. In [4], a silhouette extraction approach is developed in YUV color space that is able to

detect shadows and segment the human body from the background.

Although there are a number of algorithms developed in the literature for silhouette extraction and human tracking, most of them have targeted and work efficiently only in controlled environments, where the lighting condition is good, and the background is simple and nearly static. In this work, as a part of our project to develop automated video-based activity analysis for eldercare, we propose an efficient and robust silhouette extraction and human tracking algorithm in a real-world, unconstrained environment. More specifically, we will use video cameras to collect information about elderly residents' daily activity at home. From these videos, we extract important activity information to perform automated functional assessment and detect abnormal events, such as people falling on the floor.

To design an efficient video activity monitoring system for eldercare, there are two major issues we need to consider. The first one is privacy protection. Our approach is to extract the human body from the background and fill the corresponding image region with white pixels so as to block the identifying features. The second issue is activity analysis. Our approach is to extract feature information from the human silhouette and develop statistical models, such as Hidden Markov Models, to model and recognize activities. In [5], we proposed our baseline architecture to generate human silhouettes in dynamic temporal environments. While the results are excellent for general human motion, interaction of humans with objects in the scene still pose problems. In this paper, we address those problems through the use of fuzzy logic.

1.1 Major Challenges

In real-world environments, such as in-home eldercare monitoring, an efficient and robust silhouette extraction and human tracking algorithm has to deal with the following challenges: (1) the lighting condition is time-varying. The change in lighting could be caused by changing sunlight, passing vehicles outside the apartment, indoor lights turned on or off, adjustment of blinds, etc. (2) the background is dynamic. Objects in the background, such as furnitures and appliances are being moved by residents. (3) silhouette extraction should be accurate and robust. This is because the

subsequent vision analysis tasks, such as activity modeling and abnormal event detection, are totally based on the feature information extracted from the silhouette. This implies that the extracted silhouette should be closely conformed to the human body. For example, a simple foreground segmentation algorithm may include a chair being moved by the person as part of the silhouette. This will create a potential problem because the activity analysis module will detect a sudden change in the silhouette, from a vertical slim human body to horizontally stretched silhouette (human + chair), and may report this change as a abnormal event. According to our study, this is one of the major sources of false alarms in video-based activity monitoring.

1.2 The Baseline Approach

We formulate the silhouette extraction as an adaptive classification problem. High-level knowledge is fused with low-level feature-based classification to handle a time-varying background, and a decision process based on a fuzzy logical inference system is used to detach the moving objects from the human silhouette. Our experimental results demonstrate that the proposed silhouette extraction and human tracking scheme works efficiently and robustly.

The rest of the paper is organized as follows. Section 2 gives an overview of the baseline system described in [5]. Section 3 describes feature extraction and foreground-background segmentation. In Section 4, we develop an algorithm for adaptive background update. The decision process based on fuzzy logic inference system for detaching moving objects is presented in Section 5. Section 6 concludes the paper.

2. OVERVIEW OF THE ALGORITHM

We assume that the camera is fixed. We formulate silhouette extraction as an adaptive classification problem. The input video frame is partitioned into small blocks, for example, 4x4 blocks. For each block, a classification decision is made: the block belongs to the human silhouette or not. To this end, we extract invariant features from the image blocks, as illustrated in Fig. 1. Based on these features, we build a statistical model for the background and classify the image blocks into two categories: foreground and background. Because the background is time-varying, the background model and the classifier should be adaptive. However, background adaptive is also risky. For example, if a person sits still or sleeps in the couch for a long time, say hours, the adaptive background model will consider the person part of the background. This is not acceptable because of unprotected privacy. To solve this problem, we fuse the high-level knowledge obtained from human tracking with the low-level feature-based classification so as to guide the background update. At this stage, the foreground may still contain objects other than the human silhouette, such as objects being moved by or attached to the

person. To address this issue, we develop a decision process based on a fuzzy logic inference system to detach these objects from the human silhouette. In the following sections, we explain the proposed silhouette extraction scheme in detail.

3. FEATURE EXTRACTION

We extract features in the RGB color space. Two feature variables: chromaticity and brightness distortions are used to classify the foreground and background [6]. The color model used here separates the brightness from the chromaticity components. Let E_i be the color vector for pixel i in the RGB space and let I_i be the observed color vector. Brightness distortion α_i is defined as follows:

$$\alpha_i = \arg \min_{\alpha_i} \|I_i - \alpha_i E_i\|^2 \tag{1}$$

The chromaticity distortion is defined by

$$CD_i = \|I_i - \alpha_i E_i\| \tag{2}$$

The foreground and background classification is based on the following observation: image blocks in the background should have little change in their color distortion. The brightness distortion is very helpful in detecting shadows. In the proposed adaptive background modeling and classification scheme, we use the image data in the past Δ frames to compute the joint distribution of these to build a background model. Based on these two features and background model, we classify the image block into foreground or background. Fig. 2 shows the results of segmentation using this feature extraction and classification scheme.

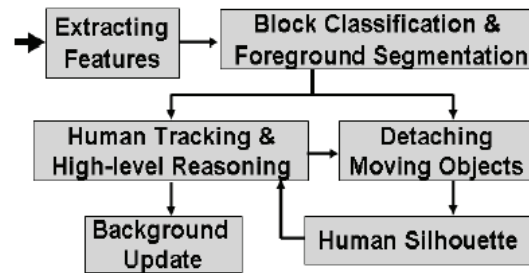


Fig.1: Block diagram of the proposed system for silhouette extraction and human tracking

4. KNOWLEDGE-BASED ADAPTIVE BACKGROUND UPDATE

It should be noted that the value of the size of the shifting window should be appropriately chosen. If it is too small, the background is updated very fast; the human body could be easily updated as background. This implies that the time duration of the shifting window should be quite large.

However, if the time duration is too large, the background is updated very slowly. If a new object is introduced into the background or a background object is moved, before the background is updated, this object will be classified into the foreground and hence become part of the silhouette. To solve this problem, we utilize high-level knowledge about human motion to guide the adaptive update of the background model.

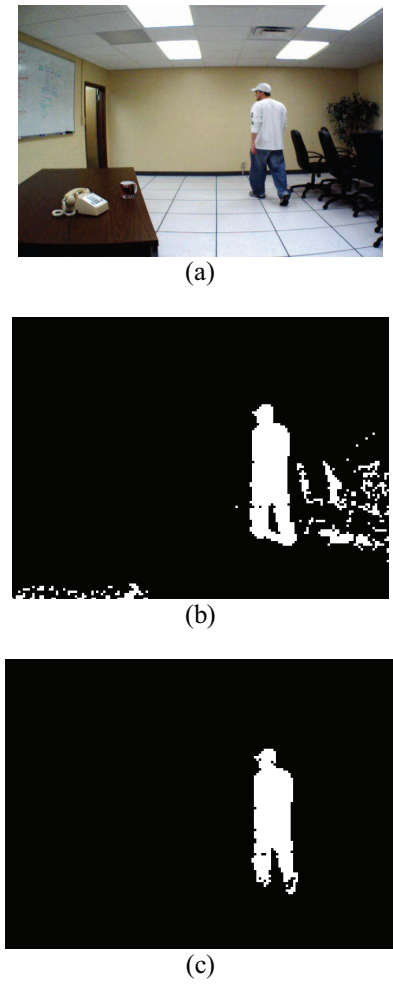


Fig.2: (a) Original image, (b) Silhouette without shadow suppression and (c) Result with shadow suppression

Our basic idea is to track the human body and predict its region in the scene. The image blocks which contain the human body should be updated very slowly such that the human body would not be updated as background. Those blocks outside the body region can be updated much faster to make sure that the new objects are quickly absorbed into the background. The prediction of human position is denoted by $V_n = (x_n, y_n, w_n, h_n)$ which define a bonding box of the human body in frame n . (x_n, y_n) and (w_n, h_n) denote the centroid and dimension of human body

silhouette in frame n separately. We use the following linear prediction model

$$V_{n+1} = \sum_{i=0}^L a_i V_{n-i} \quad (3)$$

to predict V_{n+1} for the human body in frame $n+1$. V_{n+1} then defines a bounding box, which should contain the human body if the prediction is accurate. Fig. 3 shows the human silhouette extraction with and without human movement tracking and adaptive background update.

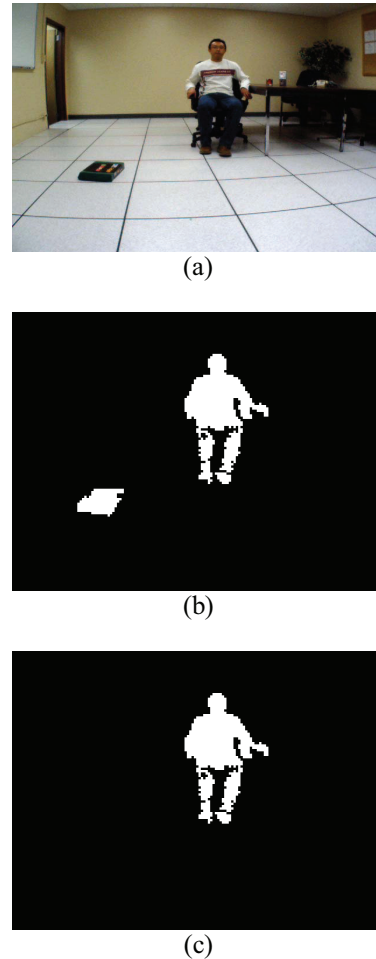


Fig.3: Background-foreground segmentation of original image (a) without (b) and with (c) human movement prediction and background update.

5. DETACHING THE MOVING OBJECTS USING FUZZY LOGIC INFERENCE SYSTEM

The main contribution of this paper is a methodology to detach objects being moved by the person from the human silhouette. Detaching objects from the human silhouette is a challenging task because of the lack of automated reasoning about what is, and is not, a human being, as evidenced in a

video sequence. Certainly, sophisticated human recognition and identification algorithms could be used. However, these algorithms are often computationally intensive and not robust in a dynamic environment like a home. Therefore, in this work, we propose a simple yet efficient algorithm for a object detachment. Our scheme is based on a fuzzy logic inference system [7], which fuses multiple sources of information together for decision making. In effect, we will create a fuzzy silhouette that we will harden by displaying an appropriate α -level cut.

Suppose we are working on frame n and the human silhouette in frame $n-1$ has been correctly extracted, denoted by O_{n-1} . Let the foreground image region in frame n be O_n , which might contain the human body and moving objects. Our fuzzy logic inference system is based on the following observations:

(1) If an image block in O_n belongs to the human body, it should have a high possibility of finding a good match in O_{n-1} . We use the SAD (sum of absolute difference) to measure the “goodness” of matching.

(2) If a lot of blocks in its neighborhood have good matches in O_{n-1} , it is a high possibility that this block also belongs to the human body.

(3) If this block is far from the predicted position of the human body centroid, as described in Section 4, it is a low possibility that this block belongs to the human body.

Based on these observations, we extract the following feature variables from each block in O_n :

- SAD (Sum of Absolute Difference) in motion matching. For each block in O_n , we find its best match in frame $n-1$. The SAD between this block and its best match form the first feature variable.
- The number of neighboring blocks having a good match in the human body.
- The distance between the new block and the predicted human body centroid.

To determine the membership functions of these feature variables, we created some ground truth for human silhouettes using several training video sequences and found their distributions, shown in Fig. 4.

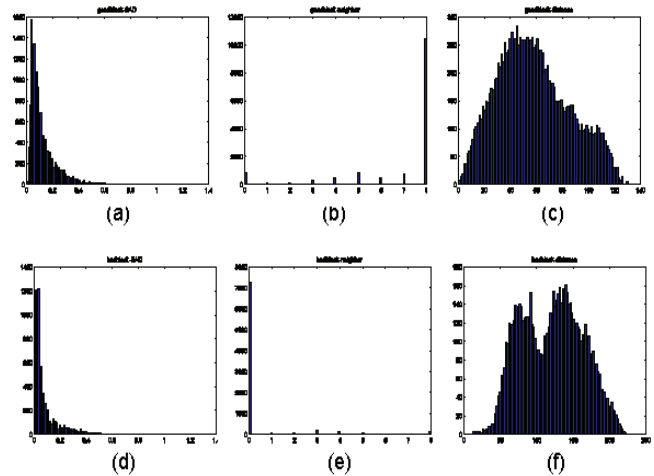


Fig.4: Distribution of three feature variables based on manually obtained ground truth, (a), (b) and (c) for correct classification, (d) (e) and (f) for mis-classification.

A set of membership functions were defined from these distributions. The following 12 rules are used in our fuzzy logic inference system, and is illustrated in Fig. 5 in a MATLAB implementation [8].

1. If SAD is very low AND Neighborhood is huge AND Distance is close, THEN Silhouette is high.
2. If SAD is large AND Neighborhood is small AND Distance is very far, THEN Silhouette is low.
3. If SAD is low AND Neighborhood is large AND Distance is medium, THEN Silhouette is high.
4. If SAD is medium AND Neighborhood is medium AND Distance is medium, THEN Silhouette is medium.
5. If SAD is large AND Neighborhood is medium AND Distance is medium, THEN Silhouette is medium.
6. If SAD is large AND Neighborhood is large AND Distance is close, THEN Silhouette is high.
7. If SAD is medium AND Neighborhood is large AND Distance is medium, THEN Silhouette is high.
8. If SAD is medium AND Neighborhood is large AND Distance is close, THEN Silhouette is high.
9. If SAD is very large AND Neighborhood is small AND Distance is far, THEN Silhouette is low.
10. If SAD is medium AND Neighborhood is small AND Distance is very far, THEN Silhouette is low.
11. If SAD is low AND Neighborhood is huge AND Distance is medium, THEN Silhouette is high.
12. If SAD is low AND Neighborhood is medium AND Distance is medium, THEN Silhouette is high.

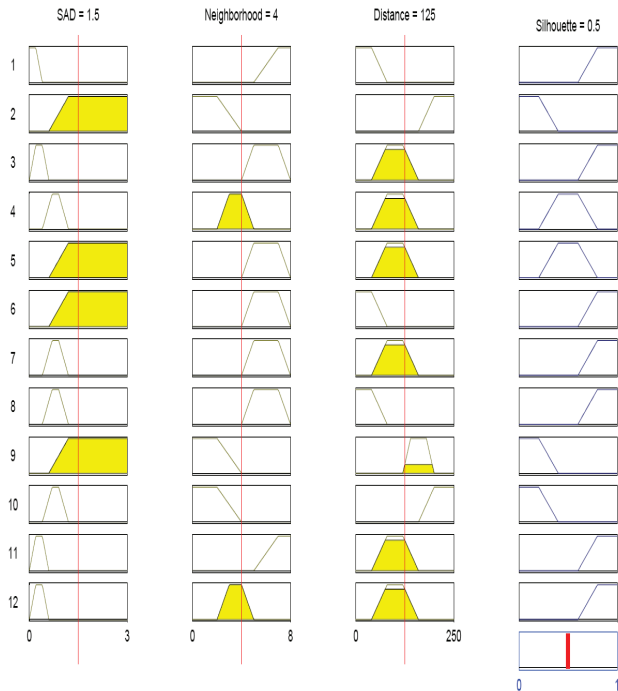


Fig.5: Membership functions of 12 rules for the fuzzy logic inference system.

We observe that using the fuzzy inference system to detach moving objects will erode the human body silhouette due to mis-classification, as shown in Fig. 6. We use morphological operations and neighborhood information to repair these missing parts.

5.1. Experimental Results

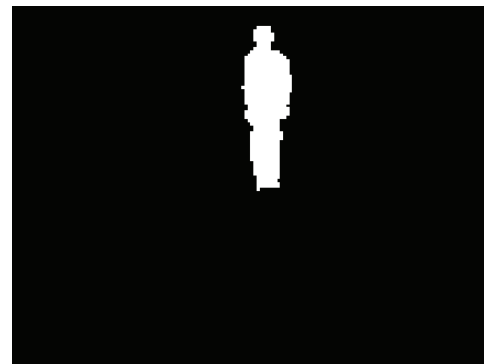
The following experimental results on several sequences show that the proposed fuzzy inference system for moving object detachment not only preserves the advantage of the low-level feature-based silhouette extraction algorithm, but also robustly detaches the moving object from the human body. As illustrated in Figs. 7 to 10, column (a) contains the original images; column (b) shows the results of feature-based silhouette extraction, i.e., with no fuzzy logic processing, and column (c) displays the results after moving object detachment using the fuzzy logic inference system. Here, the white pixels correspond to blocks in the 0.5 level cut of the fuzzy silhouette.

Fig. 7 shows the results of the sequence “Raising arms”. This sequence does not contain any moving object attached to human body. It is displayed to show that our temporal fuzzy logic processing still maintains an accurate silhouette in the baseline scenario. Columns (b) and (c) are very similar in this sequence, as they should be. Fig. 8 contains the results for the sequence “Raising a book” in which a book is attached to and moved by the person in the scene. This is a complex problem since the book is small. Single frame silhouette detection in our standard system cannot

detect that the book is not part of the human body. However, the fuzzy logic processing cleanly, clearly a superior result.



(a)



(b)



(c)

Fig.6: Degradation of the silhouette due to mis-classification: (a) the original frame; (b) the silhouette results before detaching algorithm; (c) the fuzzy silhouette. The dark gray in the edge of the silhouette is the degradation due to misclassification.

In Fig. 9, a sequence called “Falling down” in which the human has dramatic motion change is displayed. These results also verify the robustness of the proposed detaching algorithm. The fuzzy system even has the added benefit of removing some of the dark shadows that our color-based

algorithm fails to detach. However, some of the hardened silhouettes appear slightly eroded. Since the goal is to perform activity analysis, the eroded silhouettes should not cause serious problems, although we are investigating better algorithms to fill out the fuzzy shape.

Fig. 10 contains the results on one frame of the sequence “Moving chair” in which the human grabs a much bigger object. While the fuzzy logic system performs better than the corresponding crisp one, a good portion of the human silhouette is lost. As the sequence progresses, this error is compounded. Hence, we need more research on techniques to “rebuild” the displayed crisp silhouette from the fuzzy silhouette and methods to reacquire a good silhouette occasionally during the sequence. Since we will process long sequences of activity, a reasoning module should be able to detect when the human is stationary for a short time. Then, we believe that we can “reset” the golden standard silhouette.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an accurate and robust silhouette extraction scheme for a dynamic environment. We built a background model and extract classification features in RGB color space. To deal with the challenges of silhouette extraction in dynamic environment, we fused high-level knowledge and low-level features and developed a fuzzy logic inference system working on three feature variables: SAD, Neighborhood and Distance to the centroid of the human body to produce a fuzzy silhouette that was subsequently hardened by viewing an appropriate level cut. The results on several sequences show that this algorithm is efficient and robust for the dynamic environment with new objects in it. However, from results in Fig. 9 and Fig. 10, we can see that the simple morphological “cleaning” operation is not sufficient to completely recover the degradation for some sequences, especially those with greater motion and much larger objects in the scene. We are currently working on making the prediction more accurate and creating a scheme to recover missing silhouette parts using known results from feature-based classification. Also, we intend to study the impact of the accuracy of silhouette results on the performance of future activity modeling and analysis.

ACKNOWLEDGEMENTS

The authors are grateful for the support from NSF ITR grant IIS-0428420 and the U.S. Administration on Aging, under grant 90AM3013.

REFERENCES

- [1] L. Wang, T. Tan, H. Ning and W. Hu, “Silhouette Analysis-Based Gait Recognition for Human Identification”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, , pp.1505-1518, December 2003.
- [2] Y. Kuno, T. Watanabe, Y. Shimosakoda, and S. Nakagawa, “Automated Detection of Human for Visual Surveillance System”, *Proc. Int’l Conf. Pattern Recognition*, pp.865-869, 1996.
- [3] A. Elgammal, D. Harwood, L. Davis, “Non-parametric Model for Background Subtraction”, *6th European Conference on Computer Vision*. Dublin, Ireland, June/July 2000.
- [4] O. Schreer, I. Feldmann, U. Golz and P. Kauff, “Fast and Robust Shadow Detection in Videoconference Applications”, *4th EURASIP-IEEE Region 8 international Symposium on Video/Image Processing and Multimedia Communications*, pp371-375, June 2002.
- [5] X. Chen, Z. He, D. Anderson, J. Keller, and M. Skubic, “Adaptive Silhouette Extraction and Human Tracking in Complex and Dynamic Environments”, *Proceedings, International Conference on Image Processing*, Atlanta, October, 2006.
- [6] T. Horprasert, D. Harwood, L. Davis, “A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection”, *Proc. IEEE ICCV’99 FRAME-RATE Workshop*. Kerkyra, Greece, September 1999.
- [7] George J. Klir, Bo Yuan, “Fuzzy Sets and Fuzzy Logic”, Prentice Hall, 1995.
- [8] “Fuzzy Logic Toolbox User’s Guide”, The Math Works, Inc. 1998

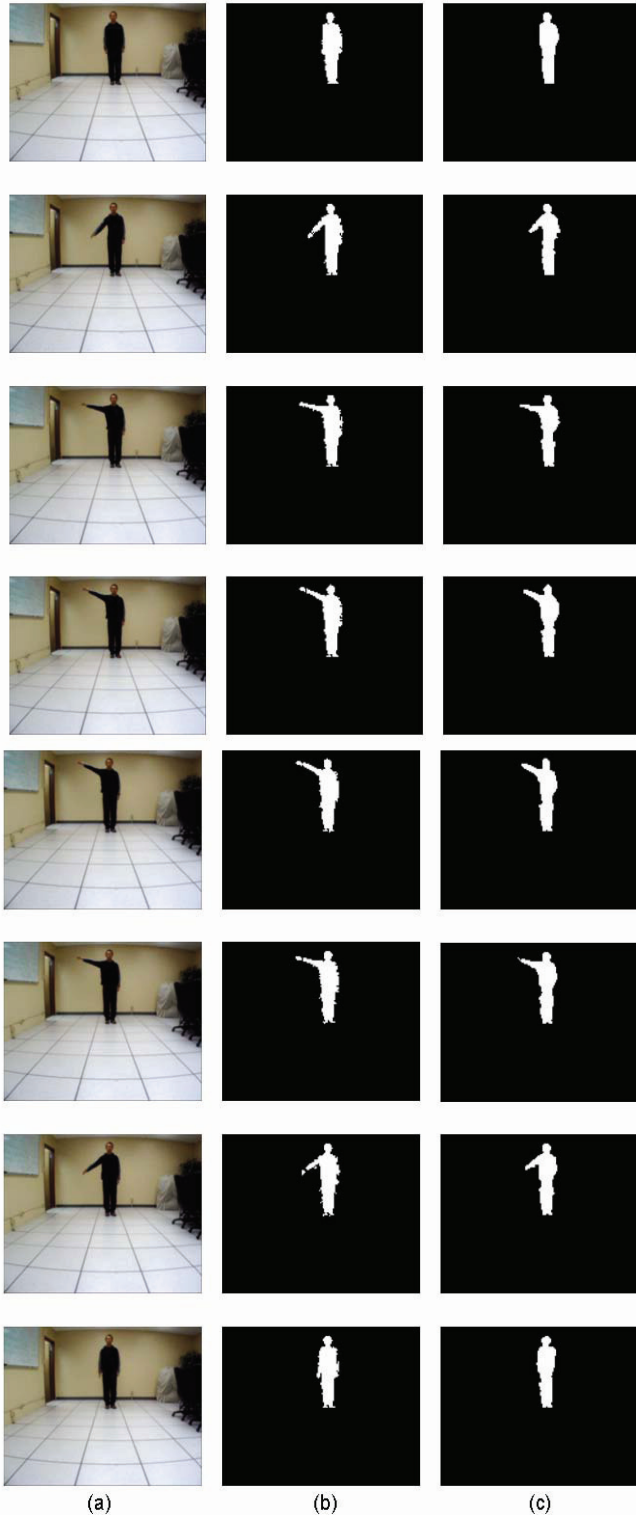


Fig. 7: The sequence “Raising arms”; (a) the original color images in the sequence, (b) the single frame silhouette extraction, (c) hardened fuzzy silhouette extraction using temporal information.

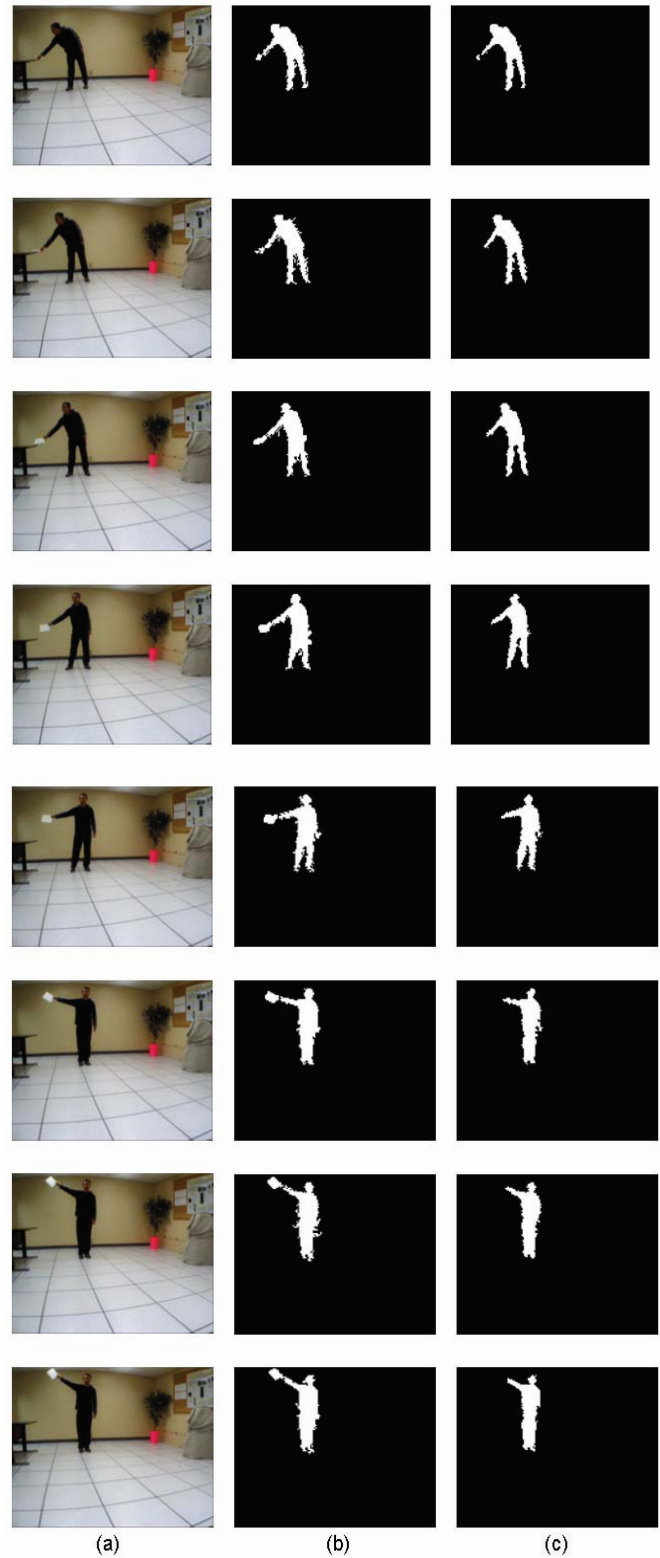


Fig. 8: The sequence “Raising a book”; (a) the original color images in the sequence, (b) the single frame silhouette extraction, (c) hardened fuzzy silhouette extraction using temporal information.

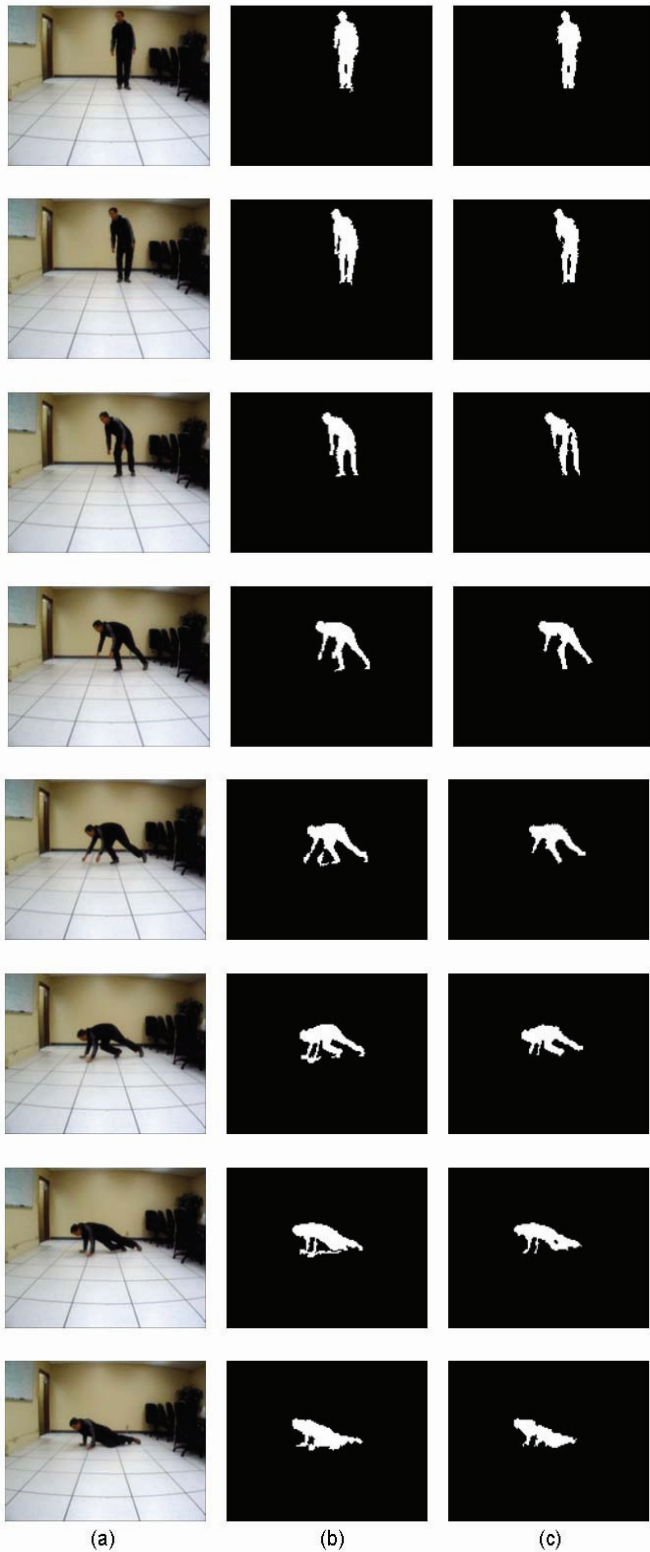


Fig. 9: The sequence “falling down”; (a) the original color images in the sequence, (b) the single frame silhouette extraction, (c) hardened fuzzy silhouette extraction using temporal information.



(a)



(b)



(c)

Fig.10: One frame from the “Moving chair” sequence; (a) the original color image, (b) the single frame silhouette extraction, (c) hardened fuzzy silhouette extraction using temporal information.