MOTION EXTRAPOLATION VIA HUMAN MOTION ANALYSIS

JOSEPH C. TSAI

Department of Computer Sci. and Information Engineering, Tamkang University, Taiwan
kkiceman@gmail.com

HUI-HUANG HSU[1], SHIN-MING CHANG[2], YING-HONG WANG[3], CHIA CHENG CHAO[4] AND
TIMOTHY K. SHIH[5]

[1,2,3]Department of Computer Sci. and Information Engineering, Tamkang University, Taiwan
[4] Information Science department of National Taipei University of Education,
[5] Department of Computer Science and Information Engineering, Asia University, Taiwan
timothykshih@gmail.com

We propose a novel motion analysis algorithm by using the mean-shift segmentation and motion estimation technique. Mean shift algorithm is frequently used to extract objects from video according to its efficiency and robustness of non-rigid object tracking. For diminishing the computational complexity in searching process, an efficient block matching algorithm: cross-diamond-hexagonal search algorithm was used. In the motion analysis procedure, the stick figure of object obtained by thinning process is treated as guidance to gather the statistics of motion information. The experimental results show that the proposed method can provide precise description of the behavior of object in several video sequences and extrapolate human motion seamlessly by combining different motion clips obtained from other video sequences.

Key words: Motion analysis, Object tracking, Object segmentation, Motion extrapolation

1   Introduction

To analyze the behavior of object in a video is an important task for object-based motion description of video. In order to analyze the detail motion information of each part of an object, techniques such as object tracking, motion estimation need to be implemented. Non-rigid object tracking and motion estimation are the most popular research topic in recent years. Among numerous tracking techniques, mean shift is the most often used. The attractiveness of mean shift algorithm [5] is due to its computational efficiency and its robustness of non-rigid object tracking. An efficient algorithm [6] using mean shift tracker and motion estimation to deal with the problem of large displacement was proposed. In addition, another paper [8] discusses moving object tracking algorithm by using mean shift-based video segmentation and edge-guided merging of over-segmented regions. The ordinary mean shift mechanism has limitation in scale space; papers [3, 6, 8] addressed the problem of scale

adaption in blob-tracking scenario and proposed an improved mean shift-based tracking algorithm to solve the problem.
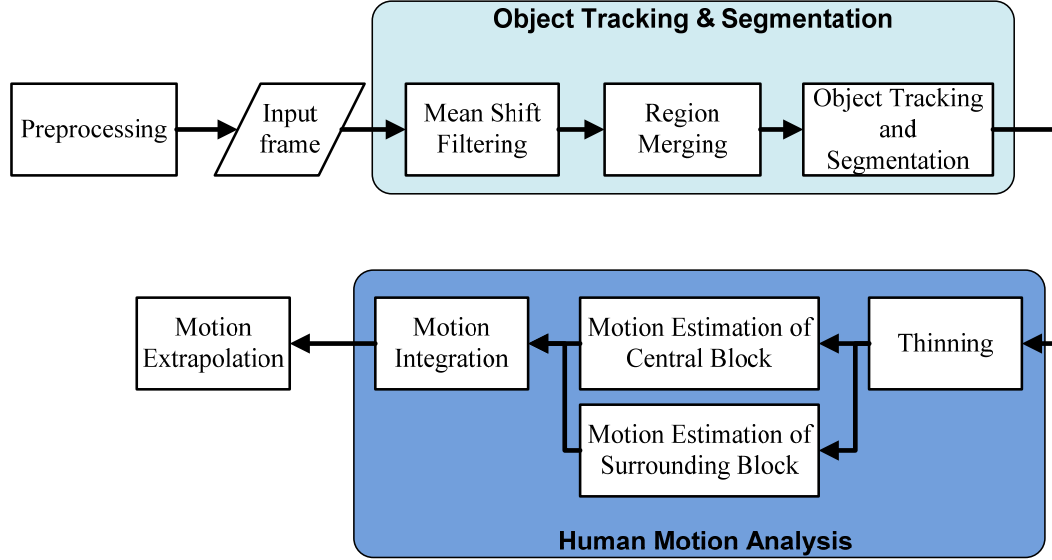

Figure 1. Overview of our motion extrapolation algorithm

On the other hand, block-matching motion estimation is one of the most important issues in video encoder. This approach is also used to enhance the accuracy of object tracking and is diminishing the computational complexity according to its similar prediction strategy. Paper [3] proposed two fast cross-diamond-hexagonal search algorithms to reduce the exhaustive searching of candidate block and maintain the accuracy of prediction.

As shown in Figure 1, our proposed method can be separated into two parts. Firstly, a modified mean shift algorithm was adopted as a color-based tracker to extract object in video frames. In this stage, we use mean shift filtering and a region-merging algorithm to produce the template of object for object tracking and segmentation process. Secondly, a thinning algorithm was adopted to generate the stick figure of object. Due to the motion information of points on stick figure can be viewed as guidance for analyzing the behavior of object, we estimating the motion vector of points in stick figure and use them as guidance to integrate the motion vector of surrounding points.

This paper is organized as the following. In Section 2, we discussed how to use mean shift segmentation and motion estimation technique to extract objects from video. The detailed description of our proposed motion extrapolation algorithm will be described in section 3, with results and conclusion in sections 4 and 5, respectively.

2   Object tracking and segmentation

As illustrated in [5], mean shift is an effective algorithm especially used in image segmentation and image smoothing. In order to extract objects in each frame from a video efficiently, a robust color

clustering algorithm and efficient estimating technique should be adopted. Our object segmentation algorithm is based on the following steps:

1. Use mean shift filtering and region-merging algorithm to build up the reference template of object.

2. The sum-of-square-difference is used to estimate the similarity and the CDHS (i.e., Cross-Diamond- Hexagonal Search) algorithm is adopted to speed up the searching process.

3. Update the reference template by extracting objects from searching result according to the template of object built in step 1.

In step 1, we use mean shift filtering technique which defined in [5] to realize color clustering. Figure 2(a) shows the original frame and Figure 2(b) is its corresponding clustering result.

(a) Source object

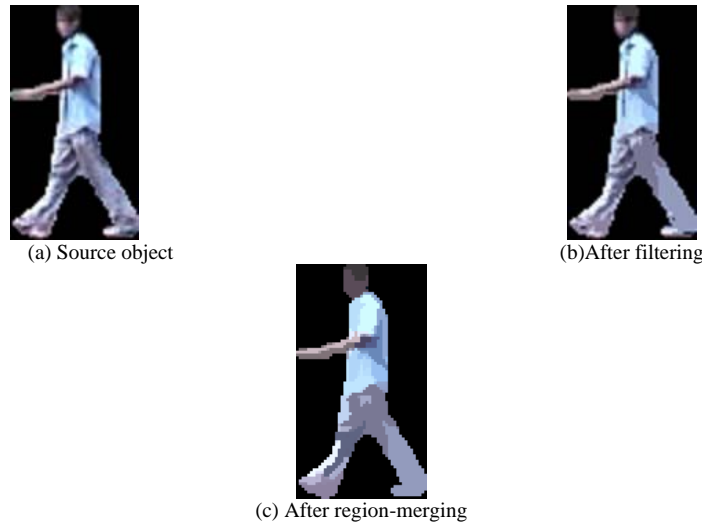(b)After filtering

(c) After region-merging

Figure 2. Color segmentation

As shown in Figure 2(b), there exist several tiny regions in the image. In order to reduce the error rate during comparing process, we use region-merging algorithm to merge regions which contains less than T_{area} pixels into surrounding larger region. T_{area} is the threshold of area and is set to 1/100 of object size. The merging result is shown in Figure 2(c) and the feature space analysis of object and segmentation is illustrated in Figure 3.

The template of object built in step 1 is used as guidance for the searching process. In our estimating procedure, the sum-of-square-difference (SSD) technique based on CIELuv color space is used to estimate similarity among the template of object and candidate patch in the next frame. In order to lower the computational cost during the estimation procedure, the CDHS algorithm is also adopted.

We used patch as a unit in our searching strategy. The sum of SSD values computed by L, U, and V (i.e., the *diff* function) could be viewed as the similarity between template of objects in previous frame and all candidate patches in the next frame.
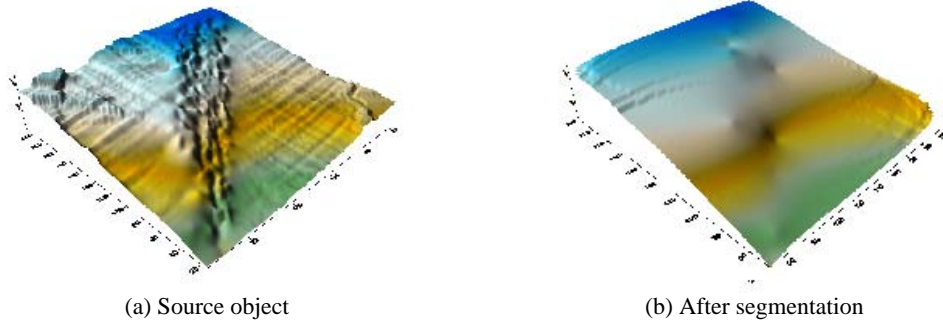

(a) Source object   (b) After segmentation

Figure 3. Feature space analysis

$$L_{SSD}(dx,dy) = \sum_{i}^{bW} \sum_{j}^{bH} \sqrt{L_t^2(x+i, y+j) - L_{t+1}^2(x+dx+i, x+dy+j)} \tag{1}$$

$$U_{SSD}(dx,dy) = \sum_{i}^{bW} \sum_{j}^{bH} \sqrt{U_t^2(x+i, y+j) - U_{t+1}^2(x+dx+i, x+dy+j)} \tag{2}$$

$$V_{SSD}(dx,dy) = \sum_{i}^{bW} \sum_{j}^{bH} \sqrt{V_t^2(x+i, y+j) - V_{t+1}^2(x+dx+i, x+dy+j)} \tag{3}$$

$$diff(P_s, P_c) = L_{SSD}(dx,dy) + U_{SSD}(dx,dy) + V_{SSD}(dx,dy), \text{ where } dx = x_s - x_c, \; dy = y_s - y_c \tag{4}$$

Let P_s be the patch which contains object's information in frame$_t$, P_c is a candidate patch in frame^{t+1}. The best searching result is a patch of the minimum difference.

$$P_{target} = \arg \min(diff(P_{source}, P_{candidate}))$$

The source patch in frame$_t$ is shown in figure 4(a) and the best searching result in frame$_{t+1}$ is shown in Figure 4(b). After step 3, the object could be extracted from the target patch according to the color clustering result of object and used in the next searching process.

3   Human Motion Estrapolation

Because of starting pose and ending pose of human motion in different motion clips are always different, for accomplishing seamlessly motion extrapolation, a procedure of human motion analysis should be adopted before we start to extrapolate the human motion. Figure 4 shows an example which we combined two different motion clips directly.

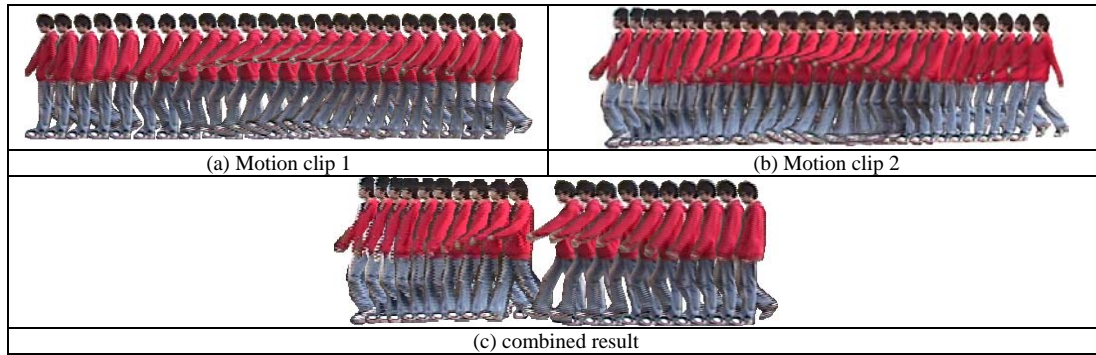| (a) Motion clip 1	(b) Motion clip 2
(c) combined result	

Figure 4. Example of motion extrapolation by direct combining two motion clips.

3.1 Human motion analysis of target objects

After objects are extracted from the above process, a thinning technique and motion estimation algorithm are used to analyze the motion information of all parts of an object.

(a) Source object	(b) Target patch	(c) Object segmentation

Figure 5. Object tracking and segmentation

In order to analyze the behavior of object, an effective thinning algorithm, Rule-Based Thinning (RBT) algorithm [1], is used to extract the skeleton of object. Points on a stick figure are viewed as the centroids of each part of object and the motion vectors of these points may present the motion of each part of object. Figure 5(a) shows the original object image and the corresponding stick figure is shown in Figure 5(b).

(a) source object	(b) thinning result

Figure 6. Result of object thinning

The motion analysis procedure is described as follows:

1. Estimate the motion vector M_c of central block (as illustrated in figure 6(a)).
2. Estimate the motion vector $M_s(i)$ of surrounding block (as illustrated in figure 6(b)).
3. Compute the mode of direction D_m in M_c and $M_s(i)$
4. If D_m is different from direction of M_c, the size of both blocks will be enlarged and return to step 1.

The block size adopted in step 1 and step 2 is 10 by 10 and 5 by 5 respectively. We also use the CDHS algorithm in the estimating process and use CIELuv-based SSD technique in the comparing process. The reason that we estimate M_c and M_r is to prevent the error occurred in estimating process and to confirm the accuracy of result. If the direction of M_c is different form D_m, we enlarge the size of a central block and surrounding block by 1 pixel and return to step 1 to estimate again.

(a) Central block   (b) Surrounding block

Figure 7. Example of central block and surrounding block in stick figure

Otherwise, M_c will be used as a reference as part of an object. Figure 7 shows the result of motion analysis procedure. Each row presents the motion vector of central block in color and direction. The first and second row shows the motion vector of central and surrounding block respectively. The result of motion integration with M_c and $M_s(i)$ is shown in the third row.

(a) Direction map of main block   (b) Direction map of surrounding block   (c) Direction map of object
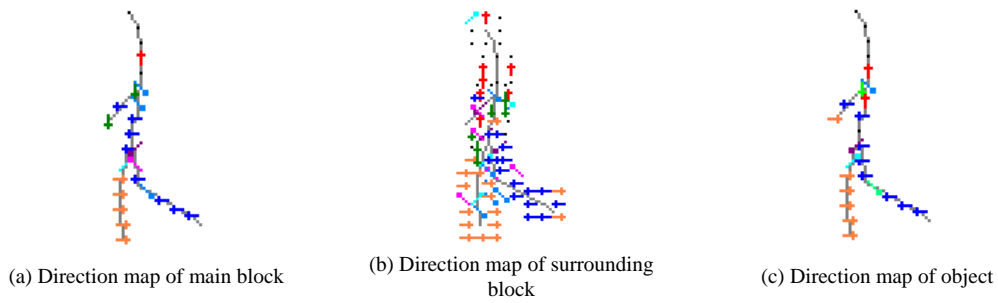
Figure 8. Example of motion analysis

Figure 8 and Figure 9 shows some results of our motion analysis procedure. Figure 8(a) and Figure 9(a) shows the original video in panning and 100-meter race in the Olympic game in front view respectively. Figure 8(b), 8(c) and Figure 9(b), 9(c) are the corresponding results of object

segmentation and motion analysis. All of computed motion information will be used as guidance in the following procedure: motion extrapolation.
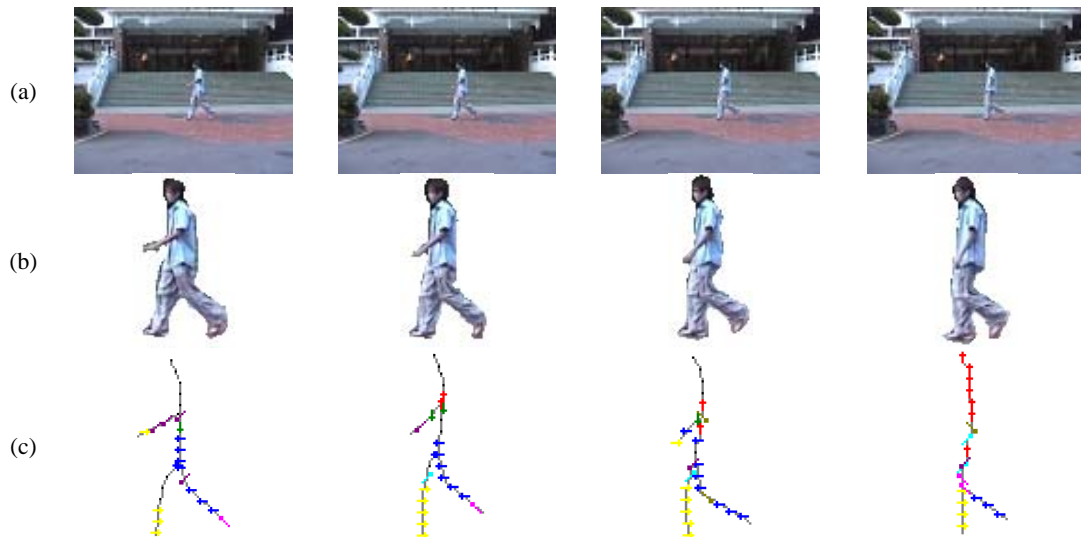
(a)

(b)

(c)

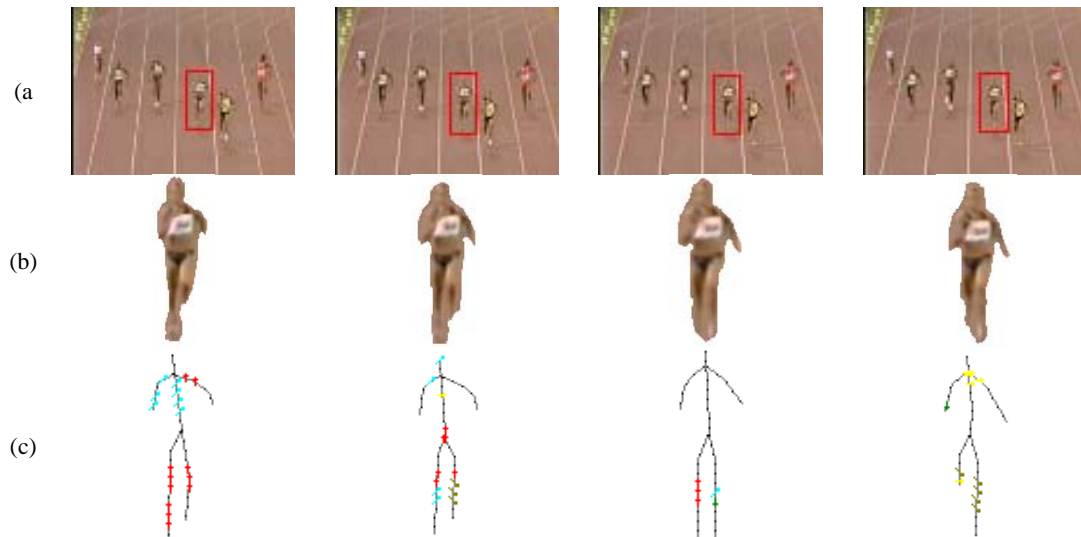Figure 9. Results of Motion analysis – Video with panning camera motion

(a

(b)

(c)

Figure 10. Results of Motion analysis –100-meter race in the Olympic game in front view

3.2 Motion Extrapolation

Since the motion information of human body is computed in previous procedure, we can start to extrapolate human motion by combining different motion clips which is obtained from other video sequences. Our motion extrapolation includes two major stages: object normalization and motion combination. The object normalization procedure is used to ensure object size in two different motion clips are the same. In this process, we use bi-linear interpolation algorithm to enlarge the smaller object. After the object size of two motion clips approximate, we can start to combine these two motion clips. In this process, we first gather the motion information of last five human motions of first motion clip and then use it as a guidance to find out most similar human motion in another motion clip. Once we can find the similar motion between two motion clips, our motion combination procedure will further combining these two motion clips. Figure 11 shows an example of motion searching and combination.

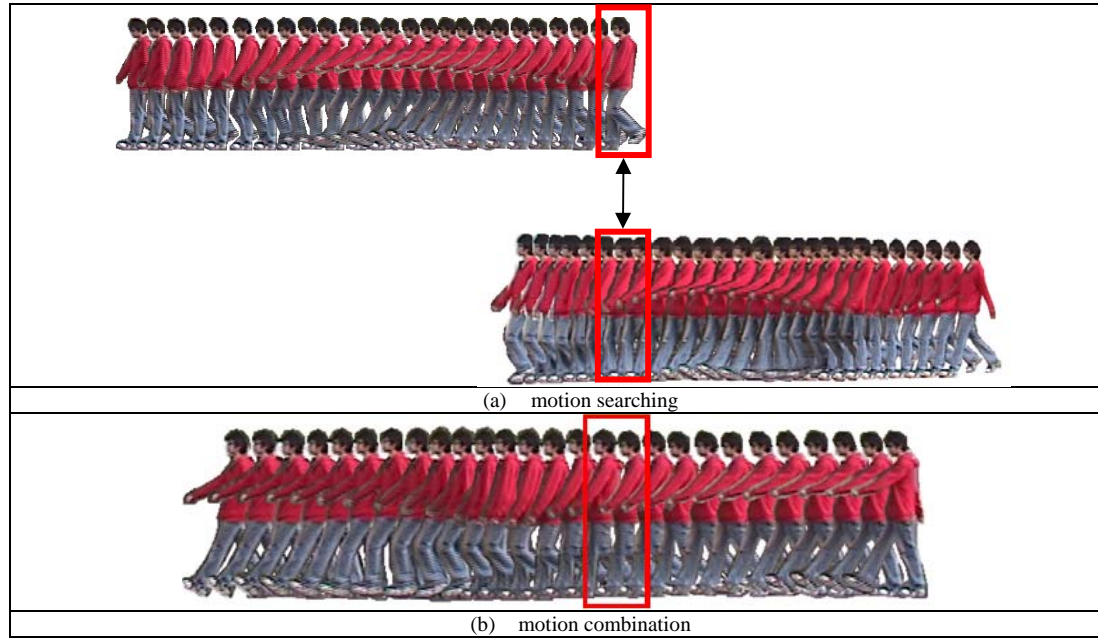(a)   motion searching

(b)   motion combination

Figure 11. Example of motion combination

4   Experimental Result

The proposed algorithm has been test on several kinds of video sequences. Figure 12 shows several results of our motion extrapolation algorithm. Each odd row is the source motion clips obtained from different video sequences. Each even row are the result of motion extrapolation. These results shows our algorithm can produce good results of motion extrapolation.
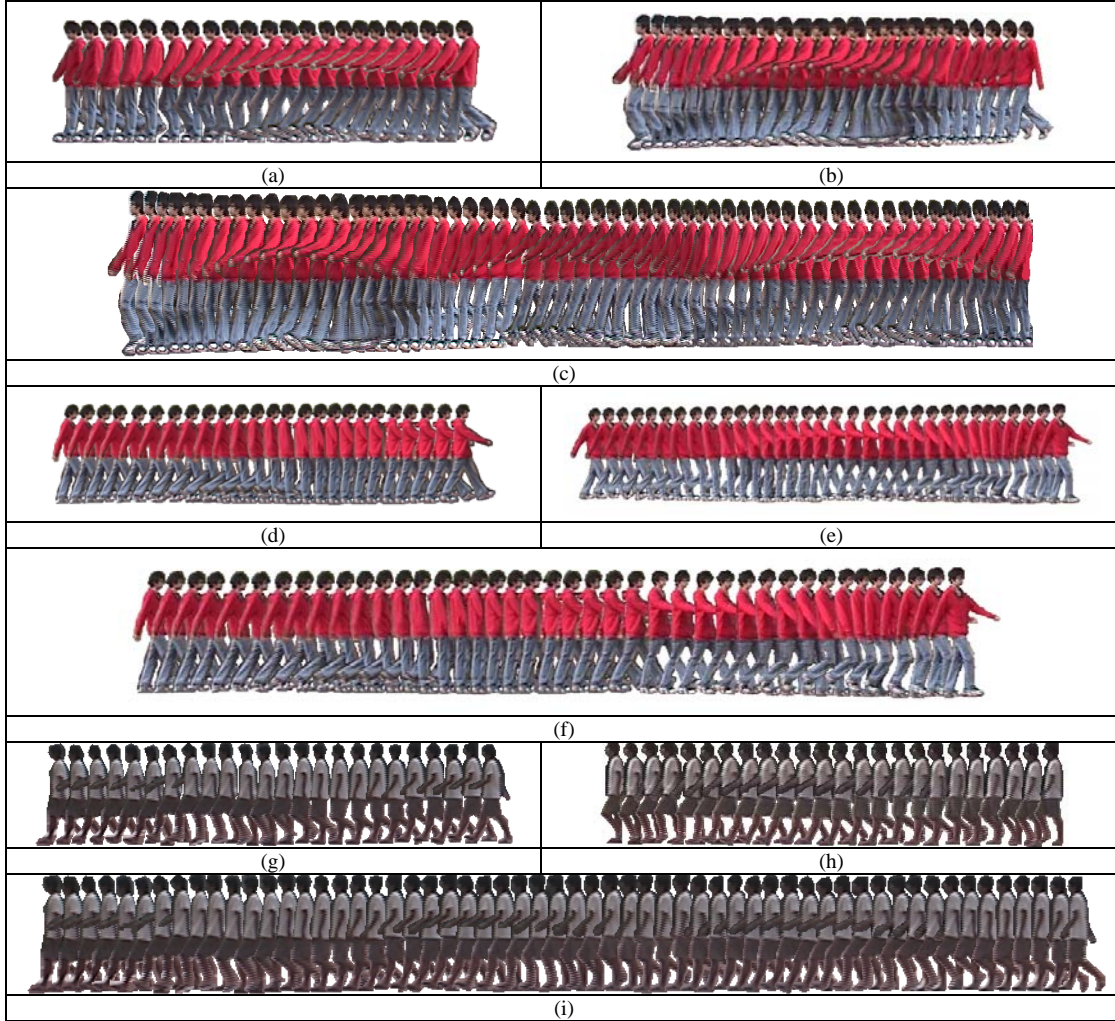
Figure 12. Results of motion extrapolation

5   Conclusion

In this paper, we proposed an effective algorithm by using the cross-diamond-hexagonal search algorithm and CLELuv-based sum-of-square-difference technique. The proposed algorithm can extract object from video and analyze the motion of each part of object precisely. Once the motion information is computed, it will be used as guidance in our motion extrapolation procedure to extrapolate human motion by combining different human motion obtained from other video sequence.

References

1. M. Ahmed, R. Ward, "A Rotation Invariant Rule-based Thinning Algorithm for Character Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, pp.1672-1678, Issue 12, 2002.
2. R.V. Babu, P. Perez, P. Bouthemy, "Robust tracking with motion estimation and kernel-based color modeling," IEEE International Conference on Image Processing, ICIP 2005., Vol.1, pp.717-720, 2005.
3. Chun-Ho Cheung, Lai-Man Po, "Novel cross-diamond-hexagonal search algorithms for fast block motion estimation," IEEE Transactions on Multimedia , Vol.7 Issue 1, pp.16-22, 2005
4. R.T. Collins, "Mean-shift blob tracking through scale space," Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.2, pp.234-240, 2003.
5. Dorin Comaniciu and Peter Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, NO.5, 2002.
6. D. Comaniciu, V. Ramesh, P. Meer, "Real-time tracking of non-rigid objects using mean shift," IEEE Conference on Computer Vision and Pattern Recognition, 2000. Proceedings., Volume 2, pp.142-149, 2000
7. Dawei Liang, Qingming Huang, Shuqiang Jiang, Hongxun Yao , Wen Gao, "Mean-Shift Blob Tracking with Adaptive Feature Selection and Scale Adaptation," IEEE International Conference on Image Processing, 2007, ICIP'07., Volume 3, pp.369-372, 2007
8. Tiesheng Wang, Irene Y.H. Gu, Andrew Backhouse, Pengfei Shi, "Moving Object Tracking from Videos Based on Enhanced Space-Time-Range Mean Shift and Motion Consistency," IEEE International Conference on Multimedia and Expo, 2007, pp.2002-2005, 2007.
9. Alper Yilmaz, "Object Tracking by Asymmetric Kernel Mean Shift with Automatic Scale and Orientation Selection," IEEE Conference on Computer Vision and Pattern Recognition, 2007, CVPR'07., pp.1-6, 2000