

Extraction of Multiple Motion Trajectories in Human Motion

Junghye Min, Jin Hyeong Park, and Rangachar Kasturi

Department of Computer Science and Engineering
The Pennsylvania State University
University Park, PA 16802-6106 USA
{min,jhpark,kasturi}@cse.psu.edu

Abstract. This paper presents a method for extracting multiple motion trajectories in human motions. We extract motion trajectories of body parts (hands and feet) using a new method based on optical flow information. This procedure is not sensitive to complicated backgrounds or color distribution of scenes. No body part model or skin color information is used in our method. We first detect Significant Motion Points (*SMPs*) and obtain motion trajectories by connecting related *SMPs* through frames using Modified Greedy Optimal Assignment (MGOA) tracker based on the distance, motion similarity, and optical flow information. We test our approach on actual ballet sequences from videos. The resulting trajectories can be used as potential features for activity recognition.

1 Introduction

Activity recognition has been attracting many researchers' interest in computer vision area because it can be applied to many video content analysis applications. Motion trajectories of body parts can be effective features for recognition of activities especially when significant body parts' movements are involved.

In this paper, we propose a new method for extracting motion trajectories of body parts. Some previous works on extracting motion trajectories are discussed first, and then an outline of our approach is presented.

1.1 Previous Works

Many approaches to locate human body parts in video sequences have been described in the literature. One of them is a body part labeling method based on appearance [11]. In this approach, body silhouettes are classified into possible postures which is followed by geometric reasoning of body parts. The location information of the body parts is temporally integrated to obtain better performance. A probabilistic model to fuse color and motion has been used to localize the body parts (hands and faces) in another approach [12].

Body parts trajectories have been used as one of the important features for action recognition. It has been the most important feature especially for gesture recognition. A state-based approach has been adopted to represent and recognize gestures [10]. Motion

trajectories are extracted from magnetic sensors and eigen templates of hand images in this method. There is the HMM-Based threshold model approach for hand gesture recognition [6]. Skin color is used to locate hands in this approach. Body parts trajectories are used also to classify gait [7]. Head, trunk, and legs are tracked to extract trajectories, and a statistical approach for modeling individual body parts with static background is used for that purpose. Motion segmentation is employed to extract motion trajectories [5]. After motion segmentation, dominant moving regions are extracted using color and geometric information. To get continuous trajectories, the affine transformations of these regions are concatenated.

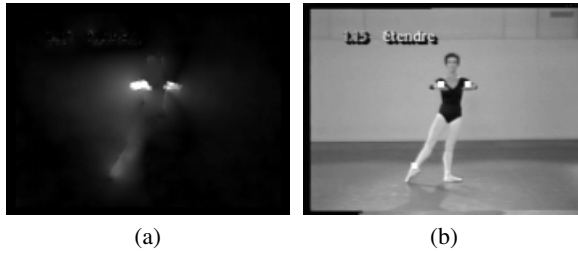


Fig. 1. (a) Optical Flow (b) Detected Significant Motion Points (SMPs)

1.2 Our Approach

In our approach, multiple motion trajectories of body parts are extracted from action sequences. To extract motion trajectories, we only use optical flow information. Many previous works depend on the skin color to locate hands and face. But skin color-based method cannot be applied to other body parts. And if background is complicated or similar colors are present in the scenes, this method cannot be stable. Human body models were also used to track body parts in previous works. These methods are affected by the type of clothing a person wears. But with optical flow, complicated backgrounds or color distribution of the scenes are not likely to cause any problem. No body model is used in this method. One other advantage of this approach using optical flow is that it can extract trajectories from the sequence with camera movement. No initialization for body parts tracking is needed which is another advantage.

After we get Significant Motion Points(SMPs) (Fig. 1) with optical flow information, we connect these points and get motion trajectories. To connect related Significant Motion Points(SMPs) and result in motion trajectories, we employed Greedy Optimal Assignment (GOA) Tracker [1] which is one of the approaches used to solve motion correspondence problem. We modified this tracker to meet our purpose.

Using the Modified Greedy Optimal Assignment Tracker, false alarms are removed and consistent movements are connected to result in motion trajectories. In Section 2, we describe how we extract motion trajectories using optical flow information. Section

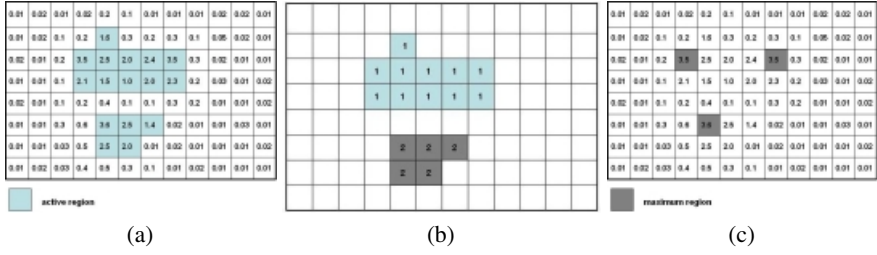


Fig. 2. (a) Active Regions (b) Labeling Active Regions (c) Maximum Regions

3 discusses the data and results of our experiments. Conclusions will be followed in Section 4.

2 Extraction of Motion Trajectories

2.1 Locating Significant Motion Points

To get optical flow from images, we used the robust gradient-based optical flow algorithm [13]. An image is divided into $N_1 \times N_2$ non-overlapping regions (Fig. 2). For each region, the average size of the optical flow is calculated. If this number is greater than the threshold value T_1 and this region is one of the top $T_2\%$ of regions which have the largest average size of optical flow, it is marked as an active region (see Fig. 2 (a)). After this procedure is applied to every region, active regions are labeled as different numbers based on the connectivity (Fig. 2 (b)). Any active region that has a greater average size of optical flow than all other active regions in neighborhoods is marked as a maximum region (Fig. 2 (c)).

For each maximum region, the point which has the largest 3x3 local average size of optical flow is found and marked as a Significant Motion Point (*SMP*) (Fig. 1(b)). But if any two *SMPs* belong to same active region based on labeled number, the distance and optical flow similarity between those points are calculated. If the distance is less than a threshold value and they have a similar optical flow direction at those points, one of these two points which has the smaller 3x3 local average size of optical flow is removed from *SMPs*.

2.2 Generating Motion Trajectories

To assign each Significant Motion Point (*SMP*) to a related trajectory, we employed Greedy Optimal Assignment (GOA) tracker [1]. Using this tracker, trajectories of fixed number of moving points can be found while allowing track continuation after point occlusion and missing detections. But track initiation and termination should be supported to generate motion trajectories of body parts. Therefore, we modified GOA tracker to support track initiation and termination as well as track continuation.

The original Greedy Optimal Assignment (GOA) tracker can be described as follows. There are M moving points, p_i , in the scene. The number of moving points, M , is fixed

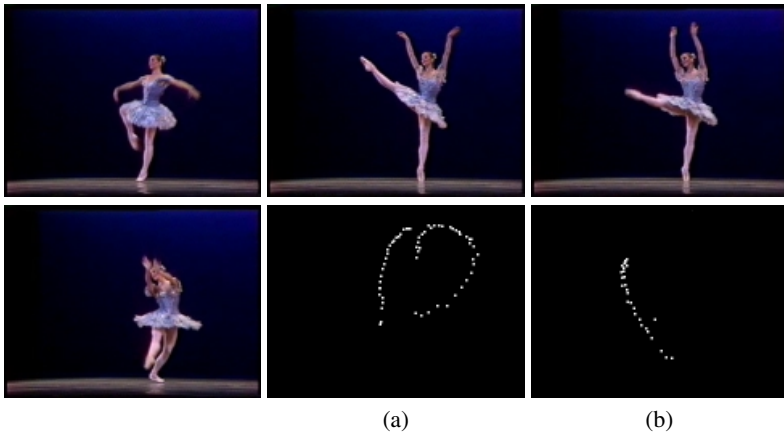


Fig. 3. The input sequence and resulting trajectories (a) two hands (b) right leg

through the sequence. Therefore, no new moving point can enter the scene, nor any moving point can leave the scene. At each frame t_k , there is a set X^k of m_k measurements x_j^k , with $1 \leq j \leq m_k$, and $1 \leq k \leq n$, of points p_i , where n is the number of frames in the sequence. The measurements x_j^k are 2D coordinates in a 2D space. The number of measurements, m_k , at t_k , can be different from M . In case of occlusion or missing detections, the m_k is smaller than M . The m_k is larger than M when there are false measurements. The objective of this tracker is to get a set of M tracks that represents the 2D motion of M moving points through the sequence. A track T_i , with $1 \leq i \leq M$, is a sequence of related measurements, $\langle x_{i_1}^1, x_{i_2}^2, \dots, x_{i_n}^n \rangle$, with $1 \leq i_k \leq m_k$. A track information up to t_k is called a track head and is denoted as T_i^k , where $1 \leq i \leq M$. To match the related track heads and new measurements at each frame, the cost of each matching is computed. Several cost functions were suggested based on the spatial proximity, motion smoothness, and the changes in the magnitude of the speed vector.

This matching problem can be described as finding an assignment matrix $A_k = [a_{ij}^k]$ with size of $M \times m_{k+1}$, where $a_{ij}^k = 1$ if and only if measurement x_j^{k+1} is assigned to track head T_i^k , otherwise, zero. The matching represented by the assignment matrix should minimize the overall cost of trajectory matching. To minimize the overall cost, the Hungarian algorithm is used [4]. The Hungarian algorithm minimizes the following expression which finds the minimal cost assignment where w_{ij} denotes the cost of assignment:

$$\text{minimize } C = \sum_{i=1}^m \sum_{j=1}^m a_{ij} w_{ij} \text{ where}$$

$$\sum_{i=1}^m a_{ij} = 1, 1 \leq j \leq m, \sum_{j=1}^m a_{ij} = 1, 1 \leq i \leq m, a_{ij} \in \{0, 1\}.$$

To apply the Hungarian algorithm, the number of elements in two sets to be matched should be equal. In Greedy Optimal Assignment (GOA) Tracker, false tracks and slave measurements are introduced to make these two numbers equal and handle the false detections and missing detections. Therefore, at each frame t_{k+1} , the assignment matrices

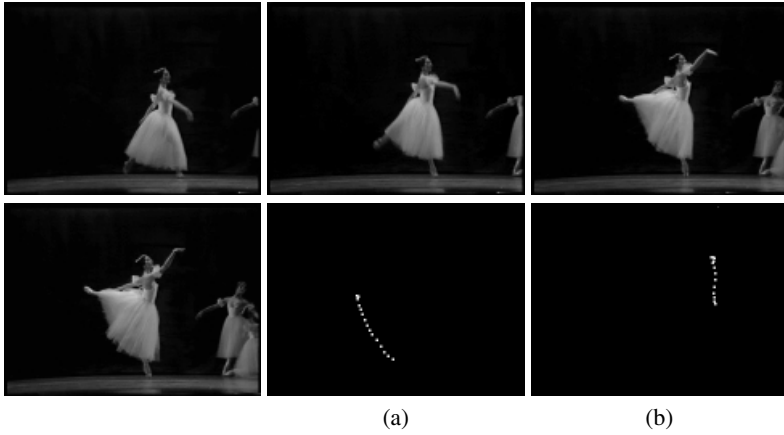


Fig. 4. An image sequence of a ballet movement and obtained motion trajectories (a) left leg (b) right hand

A^k is extended to $(M + m_{k+1}) \times (m_{k+1} + M)$ matrices A_{λ}^k by adding m_{k+1} rows and M columns, which allows for m_{k+1} false tracks and M slave measurements. These squared matrices, A_{λ}^k , enable that false measurements are assigned to false tracks and any tracks without corresponding measurements are assigned to slave measurements where each row in A_{λ}^k represents a track head and each column represents each measurement at frame, t_{k+1} .

A deviation matrix $D_{\lambda}^k = [c_{ij}^k]$ which has same size with A_{λ}^k is defined as follows. Each entry, c_{ij}^k , of D_{λ}^k is set to the assignment cost between track head T_i^k and a new measurement x_j^{k+1} in case of matching between a true track and a true measurement, i.e. $1 \leq i \leq M$ and $1 \leq j \leq m_{k+1}$. If $i > M$ or $j > m_{k+1}$, c_{ij}^k is set to ϕ_{max} , the maximum cost. If the distance between the last measurement of a track head T_i^k , $x_{i_k}^k$, and the new measurement x_j^{k+1} is greater than the maximum speed, d_{max} , c_{ij}^k is set to $\phi_{max} + \epsilon$.

With these two square matrices, the assignment problem between track heads and measurements at each frame, t_{k+1} , is solved by finding the assignment matrix A_{λ}^k which minimizes the global cost of matching where the cost of each matching is stored in D_{λ}^k . And by applying Hungarian algorithm to the deviation matrix $D_{\lambda}^k = [c_{ij}^k]$, the assignment matrix A_{λ}^k can be obtained.

In our implementation, Significant Motion Points (*SMPs*) are used as measurements, x_j^k , at each frame. To support track initiation and termination in GOA tracker and make use of the optical flow information which is available in our method, we modified GOA tracker as follows. The number of tracks, M_k , can be different at each frame, t_{k+1} , for any moving point can enter or leave the scene. Additional m_{k+1} rows in A_{λ}^k which were originally used for false tracks are used for new tracks. A new track is started from the any measurement, x_j^{k+1} , assigned to new tracks in assignment matrices, i.e. $a_{ij} = 1$ and $i > M_k$, $j \leq m_{k+1}$. For each track, the number of consecutive assignment to slave

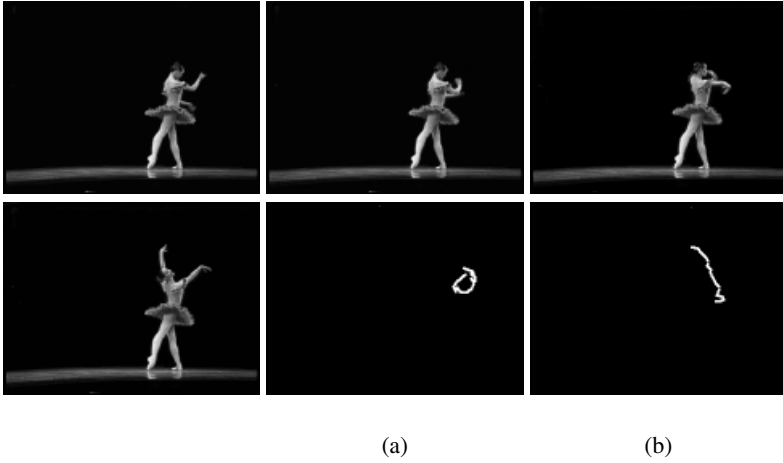


Fig. 5. The input sequence and resulting trajectories (a) right hand (b) left hand

measurements, sc_i^k , is counted at each frame, t^k . When sc_i^k is equal to a threshold value T_s , the length of trajectory up to t_k , l_i^k , is compared with another threshold value T_l . If this ratio is greater than T_l , this track is marked as a terminated track, otherwise, it is marked as a false track and removed from the list of tracks.

The number of tracks at frame t_{k+1} , M_k , is set to the number of continuing track at the frame. Therefore, the first M_k rows of A_λ^k and D_λ^k represent continuing tracks. At each frame t_{k+1} , a deviation matrix D_λ^k with size of $(M_k + m_{k+1}) \times (m_{k+1} + M_k)$ is calculated and the Hungarian algorithm is applied to this matrix, which result in an assignment matrix A_λ^k . The matching information obtained from A_λ^k is stored in each track information at each frame. If a track is assigned to a slave measurement, i.e., $a_{ij} = 1$ and $i \leq M_k$ and $j > m_{k+1}$, the calculated slave measurement is stored in track information as a new matched measurement ($x_{i_{k+1}}^k$). The slave measurement, s_j^{k+1} is calculated using optical flow information as follows.

$s_j^{k+1} = x_{i_k}^k + of_{i_k}^k$, where $of_{i_k}^k$ denotes the optical flow direction at the 2D coordinates of $x_{i_k}^k$.

The cost function is modified from the one of the cost functions used in original Greedy Optimal Assignment (GOA) Tracker. The original cost function described below penalizes changes in the direction and the magnitude of the speed vector.

$$c1_{ij}^k = 1 - \frac{(x_{i_k}^k - x_{i_{k-1}}^{k-1}) \cdot (x_j^{k+1} - x_{i_k}^k)}{\|x_{i_k}^k - x_{i_{k-1}}^{k-1}\| \|x_j^{k+1} - x_{i_k}^k\|}, c2_{ij}^k = 1 - 2 \frac{\sqrt{\|x_{i_k}^k - x_{i_{k-1}}^{k-1}\| \|x_j^{k+1} - x_{i_k}^k\|}}{\|x_{i_k}^k - x_{i_{k-1}}^{k-1}\| + \|x_j^{k+1} - x_{i_k}^k\|}$$

$$c_{ij}^k = 0.1c1_{ij}^k + 0.9c2_{ij}^k, \text{ where } 0 \leq c_{ij}^k \leq 1.$$

But in our implementation, optical flow information and the distance between the last measurement of a track and the new measurement are also considered as factors for the cost function. We introduced two more cost terms $c3$, and $c4$ which are denoted as

follows, where $c3$ penalizes the difference between the optical flow direction and the moving direction from $x_{i_k}^k$ to the new measurement.

$$c3_{ij}^k = 1 - \frac{(of_{i_k}^k) \cdot (x_j^{k+1} - x_{i_k}^k)}{\|of_{i_k}^k\| \|x_j^{k+1} - x_{i_k}^k\|}, c4_{ij}^k = \min\left\{\frac{\|x_j^{k+1} - x_{i_k}^k\|}{d_{max}}, 1\right\}$$

New cost function for matching track head, T_i^k , and the measurement, x_j^{k+1} , is defined as follows.

$c_{ij}^k = w1c1_{ij}^k + w2c2_{ij}^k + w3c3_{ij}^k + w4c4_{ij}^k$, where $w1 + w2 + w3 + w4 = 1$. At the first frame of the sequence, new tracks are started from every measurement appear in the frame. If a track is started at frame, t_k , only $c3$ and $c4$ are used as cost terms when it is matched to measurements at t_{k+1} .

3 Experiments

We used actual ballet sequences from videos (American Ballet Theater and Video Dictionary of Classical Ballet). We extracted sequences of different ballet movements. These sequences were decoded to images with a 320x240 pixel size. Each sequence has a different background and dancers with different clothing. For the cost function described in previous section, weight values, $w1$, $w2$, $w3$ and $w4$, were set to 0.6, 0.1, 0.1, and 0.2. As can be seen in figures 3- 5, motion trajectories of hand and foot are extracted simultaneously. In figure 4, there are other dancers in the background. But the trajectory of hands and legs movements are successfully extracted, which is difficult if color information was used. As can be seen in figure 5, trajectories of two hands moving across each other are extracted continuously; it is because we use optical flow information and motion smoothness for cost factors. While motion trajectories are extracted successfully in most sequences, false alarms and missed detection occur due to the shadows or objects in the background captured by moving camera. While this approach can generate motion trajectories with less limitations of background or clothing, the cost of computation time is high, mostly due to the computation of optical flow. Another limitation is that the actions without any significant body parts (hands or feet) movements cannot generate representative motion trajectories. A new way of generating whole body trajectories for the actions such as jumping or turning without significant body parts movements are also being studied.

4 Summary and Conclusions

A new method for extracting motion trajectories of body parts and its application to human action recognition is presented. Using our approach for extracting motion trajectories, multiple motion trajectories of body parts are generated without any initialization. Since no color information is used in this approach, no assumption about background is necessary. We tested our algorithms on actual ballet sequences and obtained good results. However, false alarms can occur due to shadows or camera movement.

The motion trajectories resulting from this approach can be used as potential features for action recognition and temporal segmentation of specific action in continuous video

sequences. The characteristics of body parts trajectories as features for action recognition may allow separate action recognition for each body part and this is one of the research topics we are pursuing.

Acknowledgments. The authors would like to thank Michael J. Black for making his optical flow computation code available to researchers.

References

1. C. J. Veeman, M.J.T. Reinders, E. Backer "Resolving Motion Correspondence for Densely Moving Points" *IEEE Trans. Pattern Analysis and Machine Intelligence* , Vol.23, no.1, pp. 54–72, Jan. 2001.
2. I.J. Cox, S.L. Hingorani "An Efficient Implementation of Reid's Multiple Hypothesis Tracking Algorithm and Its Evaluation for the Purpose of Visual Tracking" *IEEE Trans. Pattern Analysis and Machine Intelligence* , Vol.18, no.2, pp. 138–150, Feb. 1996.
3. C. Rasmussen, G. D. Hager "Probabilistic Data Association Methods for Tracking Complex Visual Objects" *IEEE Trans. Pattern Analysis and Machine Intelligence* , Vol.23, no.6, pp. 560–576, June. 2001.
4. H.W. Kuhn "The Hungarian Method for Solving the Assignment Problem" *Naval Research Logistics Quarterly* , Vol.2, no.1, pp. 83–97, 1955.
5. M. Yang, N. Ahuja, M. Tabb. "Extraction of 2D Motion Trajectories and Its Application to Hand Gesture Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence* , Vol.24, no.8, pp. 1061–1074, Aug. 2002.
6. H. Lee, J.H. Kim "An HMM-based threshold model approach for gesture recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence* , Vol.21, no.10, pp. 961–973, Oct. 1999.
7. D. Meyer, J. Posl, H. Niemann. "Gait Classification with HMMs for Trajectories of Body Parts Extracted by Mixture Densities," *The British Machine Vision Conference* , Vol.21,no.10 pp. 961–973, Oct. 1999.
8. J.W. Deng, H.T. Tsui "An HMM-based approach for gesture segmentation and recognition ," *In Proc. International Conference on Pattern Recognition*, Vol.3, pp. 679–682, 2000
9. J. Ben-Arie, Z. Wang, P. Pandit, S.Rajaram "Human activity recognition using multidimensional indexing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24,no.8, pp. 1091–1104, Aug.2002
10. A.F. Bobick, A.D.Wilson. "A State-Based Approach to the Representation and Recognition of Gestures," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no.12, pp. 1325–1337, Dec. 1997
11. V. Kwatra, A.F. Bobick, A.Y. Johnson "Temporal integration of multiple silhouette-based body-part hypotheses," *In Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* , vol. 2, pp. 758–764, 2001
12. E. Polat, M. Yeasin, R. Sharma "Detecting and tracking body parts of multiple people," *In Proc. International Conference on Image Processing* , vol. 1, pp. 405–408, 2001
13. M.J. Black, P. Anandan "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, 63(1), pp. 75–104, Jan. 1996.