Technical Report

---- *Yegang Tao*

Introduction

We developed a program to automatically search for the start and end points for three kinds of repeated single-action movements, including *knocking*, *lifting* and *throwing* with one of five kinds of emotions for each movement – they are afraid, anger, happiness, neutral and sadness. These motions are captured from human experimenters who were asked to perform each action with one emotion five times continuously. The positions of markers attached on body were recorded at the rate of 60Hz per frame.

For the potential research purpose, we want to split one movement into five separate movements by given the starting and ending frames for each movement, which are referred as start point S_i and end point E_i, i=1…$\#_{mv}$, where $\#_{mv}$ is the number of repeated movements in one capture sequence and $\#_{mv}$ is 5 in most cases. One could easily find S_i and E_i by visualizing the captured motion data and playback frame by frame using software like 3D-MAX, but it is a time-consuming and boring task. We would, therefore, like to propose a method to find the start and end points automatically. Since no standardized criterion to describe the starting and the ending of an action, the results produced by the program should be reviewed to check whether these start and end points are acceptable, otherwise manual correction is needed.

Method

These movements could be simply characterized from movement of the right hand[1] of the experimenter. So we work on the right hand only to obtain the start and end points. Generally, there are two basic characteristics of movement before the end points E_i: one is the velocity of right hand is decreasing continuously (negative acceleration); the other is the distance of right hand to right front shoulder is increasing continuously. Contrarily, there are two characteristics of movement after start point S_i but the velocity is increasing while the distance to shoulder is decreasing continuously. These can be used as the criterions for searching for start and end points. But these criterions are valid only for movements closed to the positions of right hand at the end point and start points. For the movements of which the positions of right hand are far from these two points, they are not sufficient conditions. This problem can be solved by removing movements far from start and end points. It produces a frame index $E_i^{'}$ of which the position of right hand is starting to approach E_i and $S_i^{'}$ of which the right hand is starting to be away from S_i. We would refer it as 'pre-thresholding' and we try to find one start and one end points within frames [$E_i^{'}$, $S_i^{'}$]. Now we can generally describe the procedure of finding the start and end points in three steps: the first step

[1] We suppose the experimenter is right-handed.

is ‘pre-thresholding’; the second step is find the end point within frames [$E_i^{'}$, $S_i^{'}$]; the third step is find the start point within frames [E_i, $S_i^{'}$].

A. Pre-thresholding

The purpose of pre-thresholding is to remove the movements of which the right hand positions are far from the start and end points, which ensure the above conditions are sufficient to find the correct start and end points. We noticed all these three kinds of movements can be roughly viewed as two parts, one is in-action part; the other is non-in-action part. Here we use a simply compare the height of right-hand with that of right front waist to determine whether a movement belong to the in-action or non-in-action part. That is:

$$\text{non}-\text{in}-\text{action}=\{\, j \mid \text{height}_j(\text{righthand}) < \text{height}_j(\text{rightfront waist}),\ j=1\ldots\#_f \,\} \qquad (2.1),$$

where $\#_f$ is the number of frames of the original movement sequence. Finally, we obtained a sequence of tuples ($E_{i'}^{'}$, $S_{i'}^{'}$), where i'=1…$\#_{nia}$ and $\#_{nia}$ is the number of segmented non-in-action movements. Frames within the range [$E_{i'}^{'}$, $S_{i'}^{'}$] are non-in-action movements. We can consider the movements in in-action part as movements far from the start and end points, and therefore try to find the i^{th} start point in [$E_{i'}^{'}$, $S_{i'}^{'}$] and i^{th} end point in [$E_{i'+1}^{'}$, $S_{i'+1}^{'}$]. Procedures of obtaining ($E_{i'}^{'}$, $S_{i'}^{'}$) are illustrated in Fig. 1.

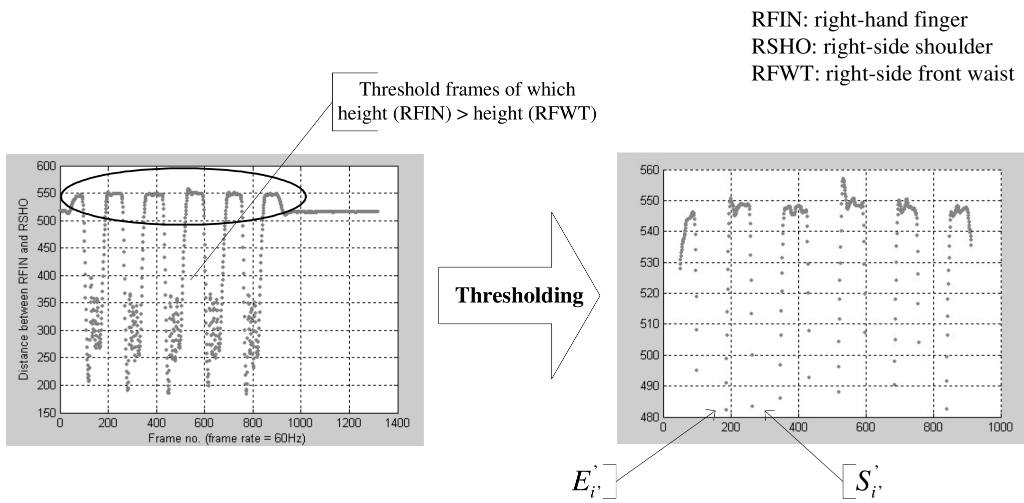


Figure 1. Pre-thresholding removes in-action movements.

Normally, thresholding produces $\#_{nia}$ = $\#_{mv}$+1 segmented non-in-action movements, but for some kinds of movements like throwing, more than $\#_{mv}$+1 segments. This is because some in-action movements are wrongly considered as non-in-action by formula (1). To remove these fake movements, we constrained $S_{i'}^{'}$ - $E_{i'}^{'}$ at least 60 (one second), since we observed that the duration of fake movements are always less than one second.

B. Find the end points

We find i^{th} end point in verified segment [$E_{i+1}^{'}$, $S_{i+1}^{'}$], except the segment [$E_1^{'}$, $S_1^{'}$] in which we can find one start point but no end point. We use two criterions and if any one of them is satisfied when we check them frame by frame the frame index E_i would be viewed as the end point of an individual movement. The first criterion c_1 is the velocity of right-hand finger is lower than a predefined threshold th_{v1}. The th_{v1} can be determined by training process. For simplicity, we observed on several movements and set th_{v1} = 5cm/s. The second one c_2 is of E_i, there is a local maximum distance (LMD) between the right-hand finger and the right-front shoulder.

We can formulate the end point of an individual movement as:

$$\text{end point} = \{f_j^{first} \mid c_1 \vee c_2\} = \{f_j^{first} \mid v_{RFIN}(f) < th_{v1} \vee f \in \{LMD_{RFIN \to RSHO}\}\} \qquad (2.2).$$

So there might be several *f*s satisfy $c_1 \vee c_2$, but we use the first one as the end point.

A1 - Algorithm of finding one end point:

a) Initialize $f = E_i^{'}$.

b) If $f \geq S_i^{'}$-1, we stop searching. $E_i^{'} = f$ and terminate.

c) If *f* satisfies (2.2), it indicates we find the end point. $E_i^{'} = f$ and terminate.

d) $f = f$+1, Go to step (b).

A Block diagram describing **A1** is given in Fig. 2.

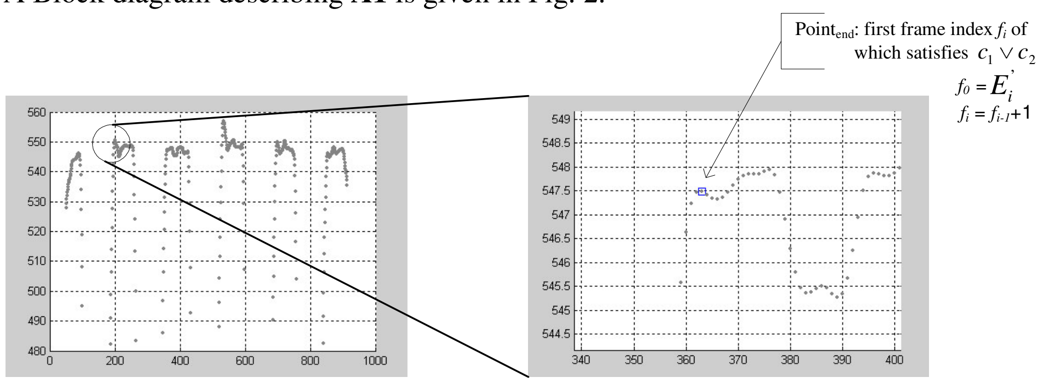


Figure 2. Block diagram of finding the end point E_i within range $[E_i^{'}, S_i^{'}]$.

C. Find the start points

After we obtained $\{E_i\}$, we find i^{th} start point in segment $[E_i, S_i^{'}]$ except the last segment $[E_{\#_{nia}}^{'}, S_{\#_{nia}}^{'}]$ in which contains no start point. For the first segment $[E_1^{'}, S_1^{'}]$, there is no end point and we use $[E_1^{'}, S_1^{'}]$ instead. The conditions to determine the start point are same to those for finding end point, but th_{v2} = 3cm/s based on observation.

We can formulate the start point of an individual movement as:

$$\text{start point} = \{f_j^{last} \mid c_1 \vee c_2\} = \{f_j^{last} \mid v_{RFIN}(f) < th_{v2} \vee f \in \{LMD_{RFIN \to RSHO}\}\} \qquad (2.3).$$

So there might be several *f*s satisfy $c_1 \vee c_2$, but we use the last one as the end point.

A2 - Algorithm of finding one start point:

a) Initialize $f = S_i^{'}$.

b) If $f \geq E_i$ +1, we stop searching. $S_i = f$ and terminate.

c) If *f* satisfies (2.3), it indicates we find the start point. $S_i = f$ and terminate.

d) $f = f$-1, Go to step (b).

A Block diagram describing **A2** is given in Fig. 3.

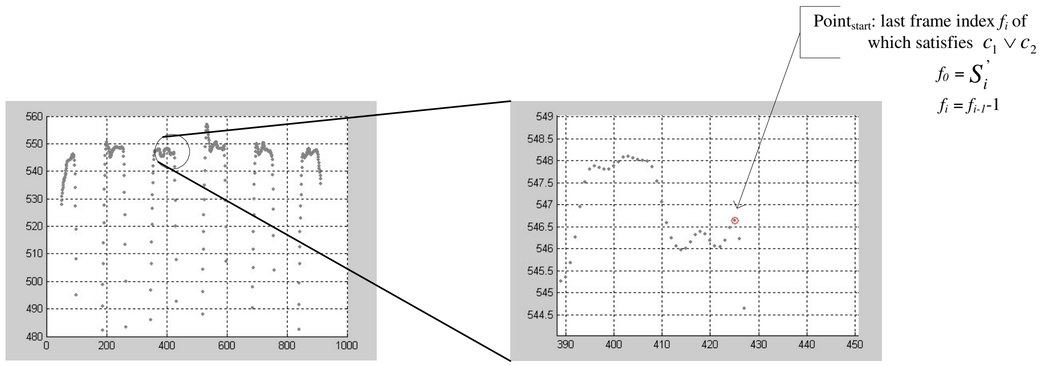


Figure 3. Block diagram of finding the start point S_i within range $[E_i\ ,\ S_i^{'}]$.

D. Post verification

Our method can always give pairs of S_i and E_i, but cannot ensure one-hundred-percent correctness. For instance, if none of c_1 or c_2 is satisfied, it will give $S_i^{'}$ and $S_i^{'}+1$ as the end and start point. Therefore, a manual verification is necessary. In 3DMAX, we let the playback of the sequential movement paused at each S_i and E_i for one second, which indicating the position of the start and end points. If they are acceptable, we will use the computer-generated results as the final results; otherwise, manually finding the start and end point is required.

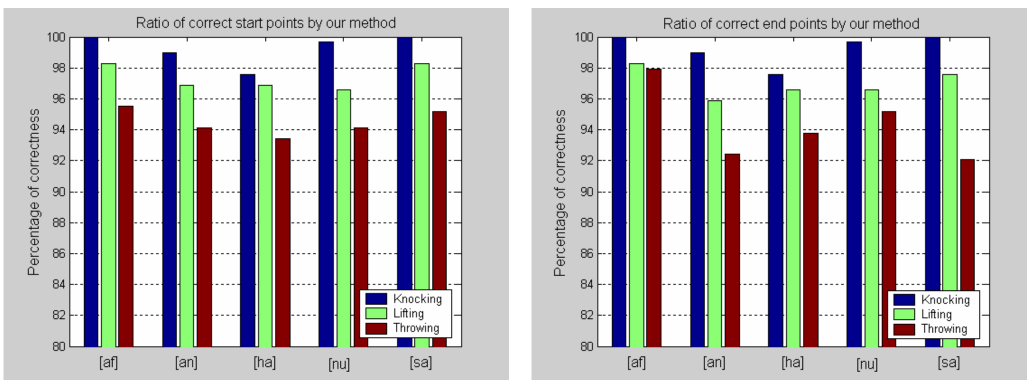
Results and Discussion

Results are obtained from 33 experimenters who were asked to perform three kinds of movements with five emotions twice – totally 2×33×3×5=990 movements. We obtain $\#_{mv}$×990 sets of { S_i^{mle} , E_i^{mle} } for each individual movement {m,e,i}, where $m \in \{k, l, t\}$ representing knocking, lifting and throwing and $e \in \{af, an, ha, nu, sa\}$ representing afraid, angry, happy, neutral and sad emotions. In Fig. 4, we show the percentage of correct start and end points obtained by our method for each kind of movement with each kind of emotion.

By comparing with different types of movements, we noticed that knocking gave the highest correctness, while throwing gave the worst results. This is because knocking movements were performed quite similar among experimenters even with different emotions; however the way of throwing involving touching object and throwing object sub-movements can be totally different person to person and can be significantly affected by emotion as well. So the general rules applied in our method worked better for small-variation movements like knocking than for large-variation ones like lifting and throwing.

By comparing the left figure the right figure, we noticed that our method is more robust to obtain the correct start points than work for end points, especially for large-variation movements like throwing. It can be explained that the way of movement for starting is simpler that that for ending, which are more consistent with the two basic characteristics we mentioned in the beginning of Chapter 2.

In post verification, we noticed and concluded three major types of movements our method cannot achieve and produced incorrect results. The first type is ‘lost’ movements. ‘Lost’ movements are some originally non-in-action movements but have too short period therefore thresholded as fake movements in pre-thresholding process; or the height of right hand are not lower than the height of right front waist and were not recognized as non-in-action movements at all. The second type is ‘extra’ movements. They are originally in-action movements but recognized as non-in-action movements and didn’t removed by fake movements’ removal process in pre-thresholding since they have more than one second duration. These two types of errors could be reduced by applying more complicated conditions to determine the in-action and non-in-action movements in pre-thresholding. The third type is ‘unfinished’ movements. In this type of errors, the end point obtained is earlier than the true end point found manually, which mostly occurs in throwing. That means from the view of right hand, it satisfies the condition for end point and seems finished, but the movements of other parts of body like shoulder and leg doesn’t return to the start position. Such errors could be removed by designing more delicate conditions for end points.



(a) (b)

Figure 4. The percentage of correct start points (a) and end points (b) for three kinds of movement with five different emotions. Percentages are averaged on $\#_{mv}$ repeated movements performed by 33 experimenters.

Conclusion

We develop an automatic method to segment continuously repeated single-action movements into separated individual movements by looking for their start and end points. The basic ideas behind our method are simple but effective. It can produce more than 90% correctness, which save the manual operations greatly. Moreover, its performance can be further improved by incorporating a training or feedback process and designing more delicate conditions for start and end points.