# Running Video Matching Algorithm Focusing on the Periodicity and Linearity of Running Motion

Yoshiteru Yamamoto[1], Keisuke Doman[1], Yoshiya Hotta[2] and Yoshito Mekada[1]
[1]*Graduate School of Engineering, Chukyo University, Japan*
[2]*Graduate School of Health and Sport Sciences, Chukyo University, Japan*

## 1 OBJECTIVES

For track-and-field runners, the running form is a major factor that affects his/her performance. The performance deeply depends on several points: the angle of an ankle when a foot touches the ground, the angle of a knee when a foot pushes the ground, the distance from the ground to the hand when swinging the arm, the speed of its swing motion and so on (Blazevich, 2010). It is, however, difficult for runners to improve their running form relying on experience and sense. Athletes and their coaches actually need an efficient way of finding running form issues for performance improvement, although they visually check their running form through performance videos.

Some useful systems using motion sensors or inertial measurement units have been developed for running form analysis (Philpott et al., 2014; Sascha et al., 2016). They, however, require large-scale devices, expensive devices or laboratory environments, which may constraint the subject's movement and then prevent the subjects from performing as usual. This research aims at supporting form analysis using cameras without any physical constraints. Note that we focus on short-distance run (100-meter) considering the significance of running form to the time record.

So far, we proposed a method for time sequence matching in order to make the running form comparison between the videos easier (Yamamoto et al., 2017). The method matched performance videos frame by frame using a dynamic time warping (DTW) framework (Myers and Rabiner, 1981) based on the similarity of running form, as shown in Fig. 1. The matching accuracy of the method should be improved for practical use, although its framework is effective. The major problem to be solved was that the method could not accurately evaluate the running form. The method calculated the running form similarity based on the runner's gait silhouette, which represents only the outline of a running form. Another problem was that the entire of an input performance video could not be used for matching due to the limitation of a DTW scheme. It is necessary that the matched videos are long enough to visually check a running form. The longer the matched video, the higher probability of finding a running form issue.

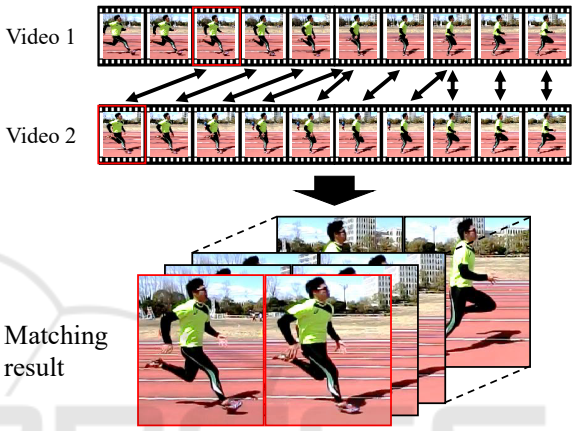


Figure 1: Time-sequence matching toward automatic running form analysis.

In this paper, we propose a more accurate and useful matching framework for time sequence matching focusing on 1) the periodicity of running motion and 2) the linearity of running speed.

## 2 METHODS

As shown in Fig. 2, running motion in general should be periodic and uniform linear motion. In the case that a performance video contains two cycles of running motion, each phase in the running motion should appear twice in the video. The proposed method matches two performance videos frame by frame focusing on the periodicity and the linearity of running motion. The method supposes that each input video contains more than one cycle of running motion and the videos are captured by fixed cameras placed at different points (50m vs. 90m).

The method is composed of four steps: 1) pose estimation, 2) correspondent phase detection, 3) linear



Yamamoto, Y., Doman, K., Hotta, Y. and Mekada, Y.
Running Video Matching Algorithm Focusing on the Periodicity and Linearity of Running Motion.
In *Extended Abstracts (icSPORTS 2018)*, pages 24-28

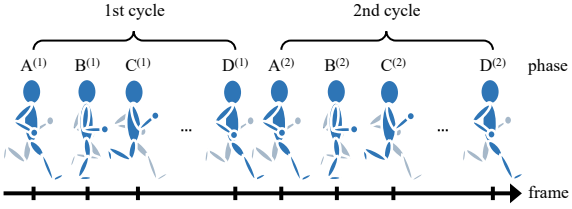


Figure 2: Periodic and uniform linear running motion.

(a) Running form at two points (b) Dissimilarity

Figure 3: Distance calculation for running form: Average Euclidean distance between body keypoints.

matching, and 4) expansion matching. The details of each step are described below.

## 2.1 Pose Estimation

The method applies OpenPose (Cao et al., 2017), a 2D pose estimation technique, to each input video, and then detects several body keypoints (joints) of the runner in each video frame. Here, OpenPose detects all of the people that appear in the videos, and does not track them. The method regards that a person with the longest hip-to-neck distance as the runner of interest for each frame.

## 2.2 Correspondent Phase Detection

### 2.2.1 Distance Measure

The method defines the dissimilarity of running form as the average Euclidean distance between the four keypoints: the left and the right wrists, and the left and the right ankles, as shown in Fig. 3. That is, the smaller the distance is, the more similar the running form is. The distance is calculated based on the coordinates of the wrists and the ankles which are represented with a neck-based relative coordinate system. This leads to ignore the scale difference of the runner's body in each frame.

### 2.2.2 Correspondent Frame Pair Detection

The method finds the correspondent timing of the running motion for each frame. For example in Fig. 2, the phase $A^{(1)}$ in the first cycle and the phase $A^{(2)}$ in the second cycle are correspondent. To do that, we introduce the following two-way detection scheme considering that the failure of pose estimation for a video

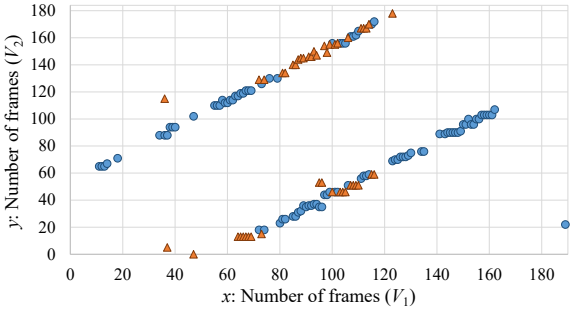


Figure 4: Example of the result of correspondent frame pair detection (The blue circle and the orange triangle indicate the pairs with the 1st and the 2nd minimum distances).

frame results in the failure of distance calculation. Let us denote each of the performance videos by $V_1$ and $V_2$. The method first finds the frame with the minimum distance $f^{(2)}_{min}$ in $V_2$ for each frame $f^{(1)}_{src}$ in $V_1$ ($V_1$ to $V_2$), and then finds the frame with the minimum distance $f^{(1)}_{min}$ in $V_1$ for each $f^{(2)}_{min}$ as a reverse check ($V_2$ to $V_1$). It is accepted if the two-way detection result agrees, that is, the frame index difference between $f^{(1)}_{src}$ and $f^{(1)}_{min}$ is less than a threshold $\theta$. Similarly, the method finds the second minimum frame pairs in $V_1$ and $V_2$. The example of a detection result is shown in Fig. 4. Note that, in this case, the two videos contain more than one and less than two cycles of running motion.

## 2.3 Linear Matching

### 2.3.1 Binary Clustering

Let $x$ and $y$ be respectively the frame indices in $V_1$ and $V_2$. The method first projects the points onto a line $y = ax$, and then clusters into two subclusters based on discriminant analysis (Otsu, 1979). Here, we uses $a = -1$ for simplicity considering that the running motion is almost a uniform linear motion and its speed does not change greatly between the input two videos.

### 2.3.2 Line Fitting with RANSAC

We can see in Fig. 4 two linear distributions with a few outliers misdetected through the two-way check. This is because the pose estimation is failed and the coordinates of misdetected body keypoints are accidentally similar. To deal with such a situation, the method applies a RANSAC algorithm (Fischler and Bolles, 1981) for robust line fitting to each subcluster. The estimated line should represent the global optimal matching between the frames in $V_1$ and $V_2$. The result of the line fitting to the data is shown in Fig. 5. The great advantage to our previous method (Yama-

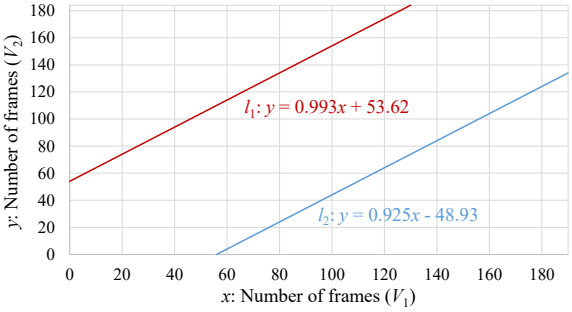


Figure 5: Result of line fitting to the data Fig. 4.

moto et al., 2017) is that the above approach is robust to the error of body keypoint detection. The method can accurately match the videos frame by frame if body keypoints are misdetected for some video frames. Note that more than one frame in $V_2$ may correspond to the frame in $V_1$ due to the multiple cycles of running motion, as shown in Fig. 5 (between the frame range of [56,130] in $V_1$).

## 2.4 Expansion Matching

Let $I$ and $J$ be respectively the total number of frames in $V_1$ and $V_2$. Let $l_1 : y = a_1x + b_1$ and $l_2 : y = a_2x + b_2$ be the two estimated lines for the subclusters 1 and 2, respectively. The method expands the video sections in both $V_1$ and $V_2$, and then matches each frame in $V_1$ with one frame in $V_2$ based on the following formula:

$$y = \begin{cases} a_2x + b_1 & (x_0 \leqq x < 0) \\ a_1x + b_1 & (0 \leqq x \leqq x_3) \\ a_2x + b_3 & (x_3 < x < I) \end{cases} \quad (1)$$

where

$$x_0 = x_1 - x_2 \quad (2)$$

$$x_1 = -\frac{b_2}{a_2} \quad (3)$$

$$x_2 = \frac{b_1 - b_2}{a_2} \quad (4)$$

$$x_3 = \frac{J - 1 - b_1}{a_1} \quad (5)$$

$$b_3 = J - 1 - a_2x_3. \quad (6)$$

As shown in Fig. 6, the method selects the frame pairs based on the three lines: $l'_2$, $l_1$ and $l''_2$ if the videos contains less than two cycles of running motion. The frames between the $x$-axis range $[x_0, 0)$ are the copy of those between the $x$-axis range $[x_1, x_2)$. Similarly, the frames after the $(J-1)$th $y$-axis index are the copy of those corresponding to the $x$-axis range $[x_3, I)$. By this expansion matching, all of the frames in both $V_1$ and $V_2$ are used for matching, which enables us to analyze the running form at all of the frames in $V_1$ and $V_2$. This is another advantage of the method.

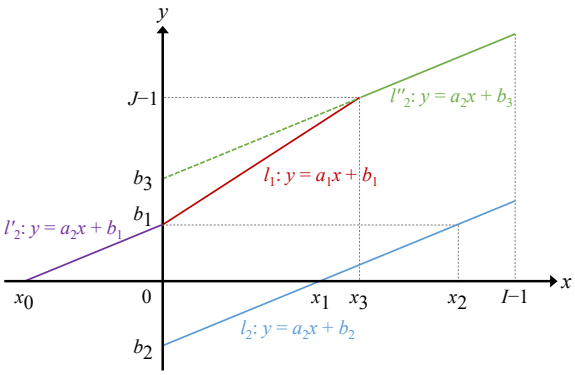


Figure 6: Expansion of the matched video section (The $x$-axis is the frame number of $V_1$ and $y$-axis is that of $V_1$).

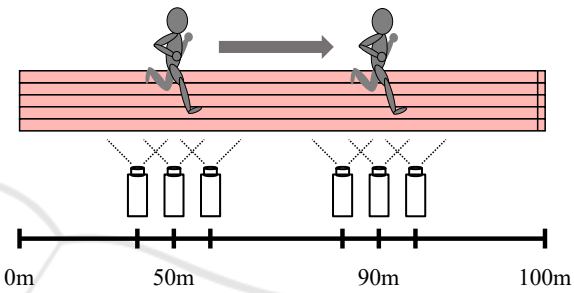


Figure 7: Camera setup in the experiment.

# 3 RESULTS

We evaluated the effectiveness and the usefulness of the proposed method through experiments.

## 3.1 Method

We captured one 100m run per runner at two points using six cameras as shown in Fig. 7. Here, the 50m and the 90m points were selected considering the maximum speed phase and the deceleration phase, respectively. The runners were five males (Runners A to E) and one female (Runner F) sprinters who were members of the track-and-field team in our university. The resolution and the framerate of the videos were 640×480 and 120 fps, respectively. We concatenated the videos captured by the three cameras into one for each point so as to contain more than one cycle of running motion. As shown in Table 1, we finally obtained six concatenated video pairs for each runner, and used them as experimental videos.

We applied the proposed method to each experimental video pair (50m vs. 90m), and then evaluated the matching accuracy based on the three measures: 1) the average matching error (the phase shift of running motion between a matched frame pair), 2) the

Table 1: Experimental videos.

| Runner | #Frames | | Cycle | |
|---|---|---|---|---|
| | 50m | 90m | 50m | 90m |
| A | 138 | 151 | 1.3 | 1.4 |
| B | 152 | 162 | 1.5 | 1.5 |
| C | 160 | 161 | 1.6 | 1.5 |
| D | 149 | 165 | 1.3 | 1.4 |
| E | 160 | 177 | 1.6 | 1.7 |
| F | 191 | 185 | 1.6 | 1.7 |

Table 2: Experimental result: Previous method (Yamamoto et al., 2017) vs. Proposed method.

| Runner | Avg. (Max.) Error | | #Matched Frames | |
|---|---|---|---|---|
| | Prev. | Proposed | Prev. | Proposed |
| A | 0.93 (4) | 0.93 (2) | 98 | 189 |
| B | 1.89 (7) | 0.68 (2) | 140 | 196 |
| C | 3.87 (9) | 1.07 (4) | 116 | 176 |
| D | 2.70 (9) | 1.41 (3) | 117 | 179 |
| E | 1.04 (6) | 1.48 (6) | 142 | 242 |
| F | 2.76 (4) | 1.69 (3) | 129 | 245 |

maximum matching error, and 3) the number of matching frames. Here, we used $\theta = 15$ for the correspondent frame pair detection considering the video framerate and the speed of running motion. Also, we compared the matching accuracy of our previous method (Yamamoto et al., 2017), which was based on the similarity of the runner's gait silhouette within a DTW-based framework (Myers and Rabiner, 1981).

### 3.2 Matching Results

Experimental results are shown in Table 2. An example of the matching result by the proposed method for each runner is shown in Fig. 8. The proposed method outperformed the previous one (Yamamoto et al., 2017) in both the average and the maximum matching error except for Runner E. Also, the greater number of matched frames were obtained by the proposed method. The coach of our track-and-field team confirmed that the matching result was accurate and enough to visual comparison of running form. The coach also commented that a seamless video containing the whole cycle of running motion helped to analyze the running motion. In this regard, such a video pair can be easily created from the matching result of the proposed method. We confirmed the effectiveness and the usefulness of the proposed method.

## 4 DISCUSSION

We discuss the effectiveness of the proposed method in terms of 1) the robustness of pose estimation error and 2) the accuracy of video matching.

### 4.1 Robustness to Pose Estimation Error

**For Runner C:** The matching error of the previous method was 3.87 (the worst among all the runners), whereas that of the proposed method was 1.07. Examples of the pose estimation results for Runner C are shown in Fig. 9. The body keypoints were sometimes misdetected, which significantly affected the running form similarity in the previous method. In contrast, the proposed method estimated the global optimal matching by line fitting with RANSAC regardless of some pose estimation failure. This was why the proposed method could achieve the accurate matching for all of the videos (runners).

### 4.2 Accuracy of Video Matching

**For Runner B:** Examples of the pose estimation results are shown in Fig. 10. The average and the maximum matching error of the proposed method for Runner B were 0.68 and 2, respectively. It was the most accurate matching among all the runners. The running motion of Runner B was more uniform and linear than that of the others at each 50m and 90m point, which is ideal for better time record. The proposed method can perfectly perform in such an ideal case. Incidentally, it may happen that the running speed changes within the range of the camera's field of view due to fatigue. We will thus study the combination with a DTW framework, and/or polygonal line fitting.

**For Runner E:** The matching error of the proposed method was larger than that of the previous one. The result of the correspondent frame pair detection for Runner E is shown in Fig. 11. No frame pairs were detected between the $x$-axis range of [0,23] due to the failure of the two-way detection. The successive lack of inliers leads to decrease the line fitting accuracy, although outliers should be ignored by RANSAC scheme. We consider that increasing the data point can solve this problem, for example, by finding not only the frame pairs of the first minimum distant frame $f^{(1)}_{min}$ but also those of the $N$-th minimum $(f^{(1)}_{min2}, f^{(1)}_{min3}, \ldots)$ for the source frame $f^{(1)}_{src}$.

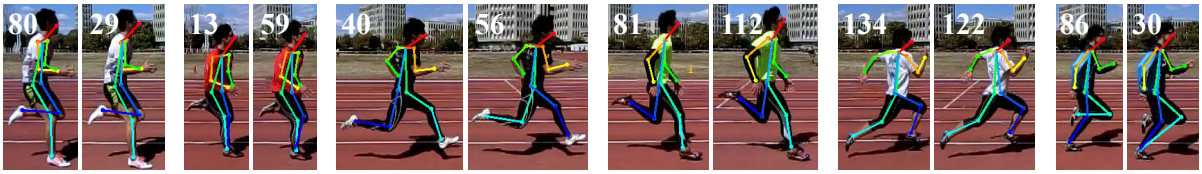


Figure 8: Examlple of the matching result for each runner (The number indicates the frame index).

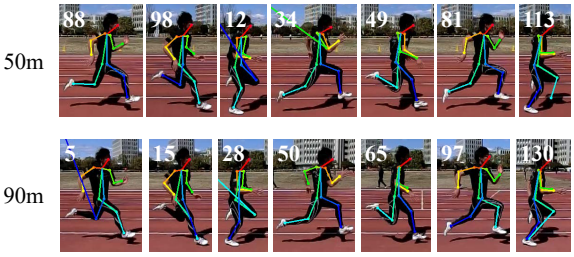


Figure 9: Examples of the pose estimation result for Runner C (The number indicates the frame index).

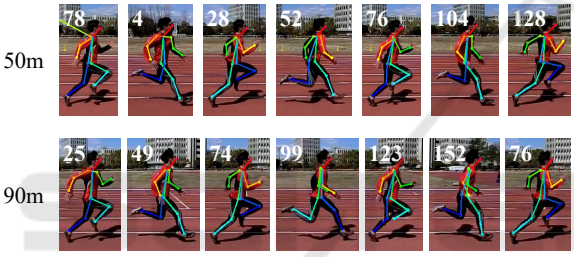


Figure 10: Examples of the pose estimation result for Runner B (The number indicates the frame index).

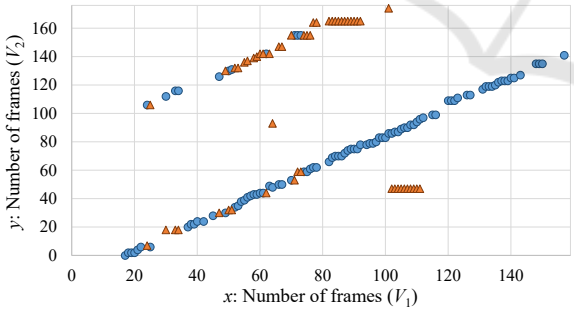


Figure 11: Result of the correspondent frame pair detection for Runner E (The blue circle and the orange triangle indicate the pairs with the 1st and the 2nd minimum distances).

## 5 CONCLUSION

This paper proposed a method for frame-by-frame video matching of running motion toward automatic running form analysis. Experimental results showed that the effectiveness and the usefulness of the proposed method. The future work includes the introduction of the combination of a DTW framework and/or polygonal line fitting, and the improvement of the correspondent frame pair detection. Also, we will develop a system that can automatically diagonose a running form issue based on matching results for performance improvement.

## REFERENCES

Blazevich, A. J. (2010). *Sports Biomechanics. The basics: Optimising human performance*. Bloomsbury Publishing, 2nd edition.

Cao, Z., Simon, T., Wei, S., and Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition*, pages 7291–7299.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. (6):381–395.

Myers, C. S. and Rabiner, L. R. (1981). A comparative study of several dynamic time-warping algorithms for connected-word recognition. (7):1389–1409.

Otsu, N. (1979). A threshold selection method from gray-level histograms. pages 62–66.

Philpott, L., Weaver, S., Gordon, D., Conway, P. P., and West, A. A. (2014). Assessing wireless inertia measurement units for monitoring athletics sprint performance. In *Proc. IEEE Sensors 2014*, pages 2199–2202.

Sascha, K., Buchecker, M., Pfusterschmied, J., Szekely, C., and Müller, E. (2016). Effects of a body-weight supporting kite on sprint running kinematics in well-trained sprinters. *J. Strength and Conditioning Research*, (1):102–108.

Yamamoto, Y., Doman, K., Hotta, Y., and Mekada, Y. (2017). Running form analysis based on time-sequense matching. In *Proc. Int. Workshop on Advanced Image Technology 2017*, number 4B-4.