Selective Use of Skeletal Information to Reduce Computational Complexity of Motion Matching

Shohei Adachi
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
alice-fr@asagi.waseda.jp

Ryohei Osawa
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
r-osawa@fuji.waseda.jp

Hiroshi Watanabe
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
hiroshi.watanabe@waseda.jp

***Abstract*—In order to improve sports skills, the comparison of the movement using the video is very effective for progress. Existing studies have been using posture similarity to map the motion timing. However, these methods have a disadvantage of a long execution time. Therefore, we propose a method to reduce the computational complexity by selectively using only data that is particularly effective among feature data used for mapping. Through evaluation experiments, we confirm that it is possible to reduce execution time without sacrificing performance even when the number of feature data is reduced. Based on the results of the evaluation experiment, we discuss the causes of the increase or decrease in accuracy for each motion and the factors that affect the accuracy of the mapping.**

Keywords*—*motion analysis, video analysis, dynamic time warping, motion matching, sport

I. Introduction

In recent years, in the field of sports, video footage of athletes during games and practice has been captured and analyzed for efficient improvement. By comparing the same action in two videos, it is easy to discover the differences between the source and the compared action in that video. This is useful for checking one's own performance and for making comparisons with skilled players.

To prevent timing deviations caused by differences in the speed of the players' movements, the conventional method uses posture similarity. Using posture similarity is effective to synchronize the timing of motion. However, while this method can estimate the skeleton from video in real time, it requires a long execution time for motion matching.

Therefore, we propose a method to reduce the execution time of motion matching by reducing the feature data used to calculate posture similarity.

II. Related work

A. OpenPose

OpenPose [1] is a method for estimating a person's skeleton based only on video data. Skeletal information can be estimated in real time, and the obtained information is output as 2D coordinate values and the confidence level of the estimation. Confidence levels are given from 0 to 1, with closer to 1 indicating greater accuracy. A confidence level of 0 means that the detection failed.

B. Start and end point free dynamic time warping

Start and end point free Dynamic Time Warping (DTW) [2] is a method that maps the elements of two time series in such a way that their similarity is maximized. When video is used as a time series, the video frames are the elements. This method calculates the difference between the elements of the two time series in a round-robin fashion and maps them so that the average value of the differences is minimized. In the mapping, all elements of the reference time series are mapped, but not all elements of the other target time series are always mapped. In addition, the target time series does not require that the start and end points of the time series coincide with the start and end points of the model time series.

C. The conventional method

The conventional method [3], denoted "Osawa method" hereafter, can provide highly accurate motion matching using posture similarity. An overview of the Osawa method is given below.

- OpenPose is used to obtain 12 skeletal information from video frames. The parts of the skeleton are the right and left shoulders, elbows, wrists, hips, knees, and ankles.
- Create spatial and temporal vector data using skeletal information. A spatial vector is the one connecting two arbitrary points selected from a skeleton of 12 points in a frame. The temporal vector is the one representing the movement of the same body part between two adjacent frames.
- The posture similarity is calculated for all frames between the model video and the target video. The difference between the spatial and temporal vectors is used as the measure.
- Match similar motion frames using the start and end point free DTW described in section II.*B*.

III. Proposed method

We propose an approach to reduce the execution time by reducing the number of spatial vectors used for dynamic time warping. In the Osawa method, all 66 vectors from the 12-point skeleton are used as spatial vectors. However, the importance of each skeleton is different depending on the type of the sport. Therefore, we propose to use only vectors that have a large influence on the matching performance as spatial vectors.

Spatial vectors, like temporal vectors, are data that represent human posture information. When the temporal vector is added as in the previous studies or the spatial vector is reduced as in our proposal, the first influence appears in the results of the posture similarity calculations. Therefore, we use Kendall's rank correlation coefficient [4] for the results of posture similarity as a measure that indicates the effect of each

TABLE I. DETAILS OF TARGET VIDEO

	Baseball video	Golf video
Number of videos	122	83
Number of average frames	172.1	340.5
Dominant hand	Right	Right

TABLE II. DETAILS OF MODEL VIDEO (PITCHING)

	Model1	Model2	Model3	Model4
Number of frames	84	80	69	77
Dominant hand	Right	Right	Left	Left

TABLE III. DETAILS OF MODEL VIDEO (SWING)

	Model1	Model2	Model3	Model4
Number of frames	59	65	48	57
Dominant hand	Right	Right	Left	Left

spatial vector on the matching performance. We first perform motion matching using only one vector out of 66 spatial vectors and investigate which vector is effective in performance. The accuracy of posture similarity calculation using only one spatial vector are sorted based on Kendall's rank correlation coefficient. Then, after considering the trade-off between performance and execution time, we choose the top 11 vectors as spatial vectors for the similarity calculation.

IV. EXPERIMENTS

A. Dataset

The effective vectors for motion matching performance are expected to vary by sports and motion types. We investigate baseball pitching and golf swing motions in the experiment. The model and target videos used in the experiment are the same as those used in the Osawa method.

The dataset contains 124 baseball videos that include a single pitching motion and 85 golf videos that include a single swing motion. We use 122 right-handed videos out of 124 baseball videos and 83 right-handed videos out of 85 golf videos as target videos. Model 1 and Model 2 are two videos each selected from these right-handed videos same as Osawa method. Model 3 and Model 4 are the remaining two left-handed videos, which are not used as target videos. The model video is clipped from the motion segment only. Details of the target video are shown in Table I, and details of the model video are shown in Tables II and III.

B. Selection of top 11 spatial vectors

To identify the effective top 11 spatial vectors, we performed motion matching using only one spatial vector for the Model 1 videos of pitching and swing. The elements of the top 11 spatial vectors, sorted by Kendall's rank correlation coefficient, are shown in Table IV and Table V. The key point 1 and 2 in the tables indicate the skeletons that make up the spatial vectors, and Kendall indicates Kendall's rank correlation coefficient. Graphical representations of the vectors shown in Tables IV and V are shown in Figures 1 and 2. In each image, the number shown in the upper left corner corresponds to the rank shown in Table IV, V.

TABLE IV. TOP 11 SPATIAL VECTORS (PITCHING)

Rank	Vector elements (kp1->kp2)		Kendall
	Key point 1	Key point 2	
1	Right Wrist	Right Ankle	0.8050
2	Left Wrist	Right Knee	0.7071
3	Right Shoulder	Left Wrist	0.7772
4	Left Wrist	Right Hip	0.7651
5	Left Elbow	Right Hip	0.7645
6	Left Wrist	Left Hip	0.7642
7	Right Ankle	Left Hip	0.7476
8	Left Shoulder	Right Hip	0.7474
9	Left Shoulder	Left Wrist	0.7448
10	Right Hip	Left Hip	0.7290
11	Left Elbow	Left Hip	0.7277

TABLE V. TOP 11 SPATIAL VECTORS (SWING)

Rank	Vector elements (kp1->kp2)		Kendall
	Key point 1	Key point 2	
1	Right Hip	Right Ankle	0.8521
2	Right Knee	Left Ankle	0.8485
3	Right Hip	Left Knee	0.8474
4	Right Knee	Right Ankle	0.8450
5	Right Knee	Left Knee	0.8358
6	Right Wrist	Left Ankle	0.8190
7	Right Elbow	Right Ankle	0.8084
8	Right Shoulder	Right Ankle	0.8015
9	Right Wrist	Left Knee	0.8007
10	Right Wrist	Left Hip	0.7947
11	Right Hip	Left Hip	0.7937

C. Motion maching by selected top 11 spatial vectors

Motion matching is performed on Models 1, 2, 3, and 4 using the 11 spatial vectors given by Table IV and V.

We use Kendall's rank correlation coefficient as an evaluation of the posture similarity calculation, as described in the proposed method in section III. In addition, as in the Osawa method, the average error frames at the start and end of the motion segment are used as an evaluation index to evaluate the mapping. This is the average of the difference between the motion start and end frames of the target video obtained by the proposed method and the motion start and end frames obtained visually. The smaller the average error frame, the more accurate the mapping results.

The execution time and performance of posture similarity calculation indicated by Kendall's rank correlation coefficient for pitching and swing are shown in Table VI and Table VII. The results for the average error frame, which indicates the accuracy of the matching, are shown in Table VIII and Table IX.

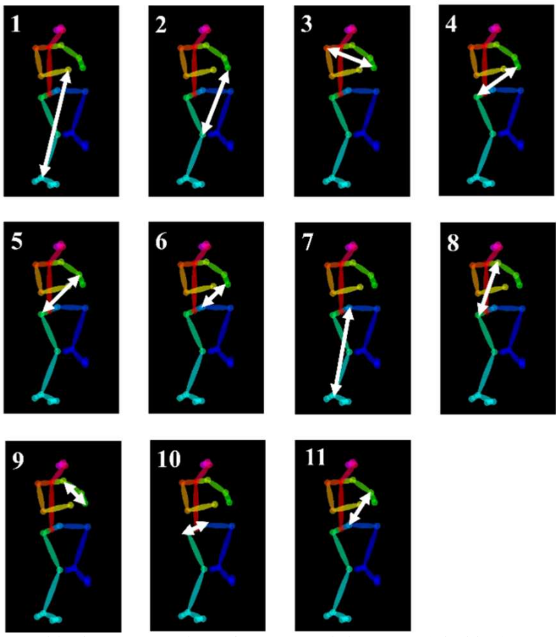

Fig. 1. Graphical representation of top11 spatial vectors (Pitching)

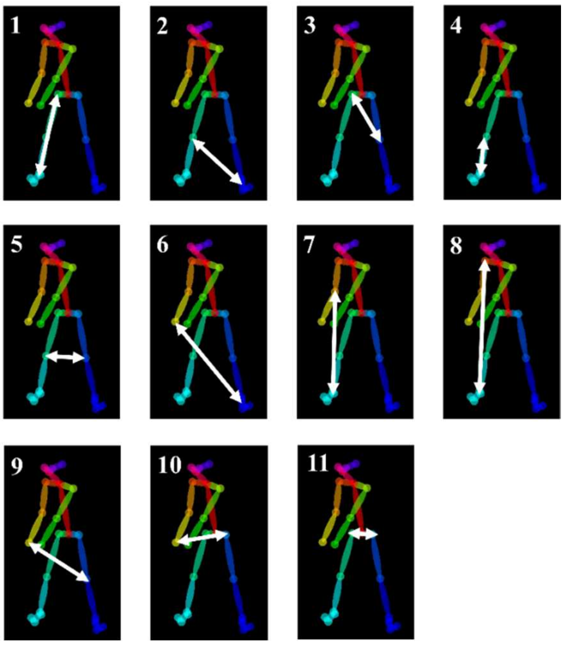

Fig. 2. Graphical representation of top11 spatial vectors (Swing)

For the pitching motion shown in Table VI, the execution time is reduced -25% without significant performance degradation for Model 1, which is used as the criterion for selecting the effective 11 spatial vectors. However, the performance of the other model videos is degraded in the range of 0.093 to 0.1405. On the other hand, for the swing motion shown in Table VII, there is no performance degradation for all model videos, resulting in improved performance and execution time is reduced from 15.6% to 26.7%.

For the pitching motion shown in Table VIII, a decrease in accuracy compared to the conventional method was observed in the average frame error as well as in the evaluation of the posture similarity. On the other hand, the average frame errors for the swing motion mapping shown in Table IX show that the accuracy was improved in some cases compared to the conventional method.

TABLE VI. POSTURE SIMILARITY EVALUATION AND EXECUTION TIME (PITCHING)

Method	**Kendall/ Time [s]**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Osawa	Kendall	0.8746	0.8192	0.8267	0.8014
	Time [s]	152.7	382.9	306.6	346.9
Proposed	Kendall	0.8672 (-0.0074)	0.7262 (-0.093)	0.7183 (-0.1084)	0.6609 (-0.1405)
	Time [s]	114.0 **(-25%)**	319.7 **(-16%)**	284.8 **(-7%)**	294.9 **(-15%)**

TABLE VII. POSTURE SIMILARITY EVALUATION AND EXECUTION TIME (SWING)

Method	**Kendall/ Time [s]**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Osawa	Kendall	0.8817	0.8966	0.9109	0.9176
	Time [s]	611.1	715.0	455.4	641.6
Proposed	Kendall	0.9013 (+0.0313)	0.9270 (+0.0304)	0.9316 (+0.0207)	0.9209 (+0.0033)
	Time [s]	495.7 **(-18.8%)**	603.3 **(-15.6%)**	353.1 **(-22.5%)**	470.6 **(-26.7%)**

TABLE VIII. FRAME ERRORS (PITCHING)

Method	**Start or End**	**Average of frame errors**			
		Model1	***Model2***	***Model3***	***Model4***
Osawa	Start	**6.36**	**4.77**	**7.21**	**5.05**
	End	**7.04**	**8.27**	**6.75**	**4.92**
Proposed	Start	8.20	13.15	19.79	14.00
	End	8.16	8.30	7.44	5.54

TABLE IX. FRAME ERRORS (SWING)

Method	**Start or End**	**Average of frame errors**			
		Model1	***Model2***	***Model3***	***Model4***
Osawa	Start	4.94	**3.95**	**3.36**	6.71
	End	7.57	**6.33**	**9.28**	**10.78**
Proposed	Start	**4.12**	4.79	3.57	**5.86**
	End	**7.50**	7.27	10.61	11.89

V. DISUCUSSION

We have two discussions of the experimental results in section IV. The first is why the accuracy of the swing motion is improved while the accuracy of the pitching motion is decreased in the evaluation by Kendall's rank correlation coefficient shown in Tables VI and VII. The second is what factors of each vector affect the influence on the accuracy of the posture similarity calculation.

A. Why the pitching accuracy decreased and the swing accuracy increased?

We have a hypothesis as to why the accuracy of the swing motion was improved while the accuracy of the pitching motion was decreased in the proposed method. The idea is that in the pitching motion, most of the 66 spatial vectors make an important contribution to the calculation of posture similarity, while in the swing motion, some of the vectors have a positive influence on the similarity calculation, while others may cause a loss of accuracy. In the pitching motion, the entire body, including the hands and feet, moves greatly in a series of motions, whereas in the swing motion, the hands move greatly

TABLE X. DISPERSION IN KENDALL'S RANK CORRELATION COEFFICIENT

	Pitching	**Swing**
Variance	0.0045	0.0103
Standard deviation	0.0671	0.1014

TABLE XI. CORRELATION BETWEEN VECTOR MOVEMENT AND ACCURACY

	ρ (Pitching)	**ρ (Swing)**
Length	-0.099	-0.069
Angle	-0.162	0.085
Length+Angle	-0.148	-0.071

while the feet are fixed to the ground and do not move much. It is possible that these differences in the features of each motion may affect the distribution of effective vectors.

Therefore, we examined the dispersion of Kendall's rank correlation coefficients for 66 posture similarity calculations using a single vector for each motion. If the dispersion of the evaluation is large, it means that the difference between effective and ineffective vectors is large among the 66 spatial vectors. When the dispersion is small, it means that the 66 spatial vectors are equally influencing the similarity calculation. Table X shows the dispersion of Kendall's rank correlation coefficients for each motion, with the variance shown in the first row and the standard deviation in the second row.

The results in Table X show that the dispersion of Kendall's rank correlation coefficient is greater for the swing motion than for the pitching motion. Compared to the swing motion, the pitching motion has 66 spatial vectors that equally affect the accuracy, resulting in lower accuracy for the proposed method that uses only the top vectors. On the other hand, the difference in the influence of the superior and inferior spatial vectors is larger in the swing motion than in the pitching motion. As a result, the accuracy of swing motion improved in proposed method, which uses only the superior vectors.

B. *What affects the accuracy of single-vector matching?*

The accuracy of the posture similarity calculation when using only the single vector shown in Tables IV and V and the results of the examination of the hypotheses made in previous section indicate that the influence of each spatial vector on the accuracy varies depending on the motion. In this proposal, we manually investigated the vectors that have the greatest influence on the pitching and swing motion. However, it is necessary to investigate what determines the influence of each spatial vector on accuracy in order to apply the proposed method automatically to various motions. Therefore, we assumed that the magnitude of the movement of the spatial vector determines its effect on accuracy, and conducted an investigation.

First, the average moving distance and the average angular change of each spatial vector are calculated in the model video1 of the pitching and swing motion. The average moving distance is the sum of the moving distance between two consecutive frames in the video for each vector, divided by the number of video frames. The average angular change is the sum of the angular change between two consecutive frames of the video for each vector, divided by the number of video frames. The spatial vectors are then sorted and ranked in ascending order based on the average moving distance, the average angular change, and the sum of the average moving distance and the average angular change, respectively. We investigate whether there is a correlation between the magnitude of spatial vector movement and accuracy by calculating the correlation between these rankings and the ranking of accuracy based on Kendall's rank correlation coefficients.

Spearman's rank correlation coefficient [5] is used to calculate the correlation. Spearman's rank correlation coefficient is expressed as ρ and takes the range $-1 \leq \rho \leq 1$. The closer ρ is to -1, the more negative the correlation is, the closer it is to 1, the more positive the correlation is, and the closer it is to 0, the more uncorrelated it is. The correlation between the magnitude of vector movement and accuracy for each motion is shown in Table XI. The first row shows the correlation between the average moving distance and the accuracy, the second row shows the correlation between the average angle change and the accuracy, and the third row shows the correlation between the sum of the average moving distance and the average angle change and the accuracy.

The results in Table XI show that the correlations ρ between the magnitude of vector movement and Kendall's rank correlation coefficients are all close to 0 for each motion, indicating that there is no correlation. Therefore, it was confirmed that at least the magnitude of the vector movement is not a factor that affects the influence on accuracy of the posture similarity calculation.

VI. Conclusion

In this paper, we propose a method for reducing the number of spatial vectors used in motion matching based on posture similarity. We selectively used only those spatial vectors that were particularly effective in calculating posture similarity. Through the simulation, we confirm that by keeping only effective spatial vectors, execution time can be reduced in the range of 15.6% to 26.7% without significant performance degradation. Based on the results of our evaluation experiments, we also discuss the causes of the increase or decrease in accuracy of each motion and the automatic selection of effective vectors.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP20K11344.

References

[1] Z. Cao, T. Simon, S. Wei and Y. Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinitiy Fields,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1302-1310, Jul. 2017.

[2] Shinya Yokoi, “Alignment Method for Body Parts Coordinates Obtained from Sport Video,” Master thesis of Waseda University School of Fundamental Science and Engineering Department of Computer Science and Communications Engineering, Feb. 2019. (in Japanese)

[3] R. Osawa and H. Watanabe “Sports Motion Matching Based on Atheletes’ Pose Similarity,” Information Processing Society of Japan (IPSJ), Technical Report of Audio Visual and Multimedia information processing (AVM) 116, No.5, pp.1-3, Feb. 2022. (in Japanese)

[4] Maurice George Kendall, “A New Measure of Rank Correlation,” Biometrika Vol.30 No. 1/2, pp. 81-93, Jun. 1938.

[5] C. Spearman, “The Proof and Measurement of Association between Two Things,” The American Journal of Psychology, Volume15, No.1, pp72-101, Jan. 1904.