Ball Motion State and Abrupt Pose Features based Player Qualitative Action Recognition for Volleyball Game Analysis

Yang Liu, Xina Cheng, Takeshi Ikenaga

Graduate School of Information, Production and Systems, Waseda University,
Kitakyushu 808-0135, Japan

leon@fuji.waseda.jp

Abstract

Volleyball video analysis is important for developing applications such as player evaluation system or tactic analysis system. Among its different topics, player action recognition is the key part for understanding player's behavior. Most existing research focused on the discrimination between different actions. The quality of an action has received little attention so far, even though the action quality potentially provides lots of useful information for volleyball analysis. Most action recognition works cannot work well due to the high similarity in different qualitative actions and appearance variation caused by target human change. This paper proposed a ball motion state and abrupt pose features based qualitative action recognition for volleyball player. Ball motion state feature is to evaluate the action quality through the ball transition caused by the action, which evaluates the return ball quality and indicates the motion quality. Abrupt pose feature is to represent the abrupt shape difference among different qualitative actions, which overcomes the high similarity in different qualitative actions. Experiments are conducted on game videos from the Semifinal and Final Game of 2014 Japan Inter High School Games of Men's Volleyball in Tokyo Metropolitan Gymnasium. The experiments show the result accuracy achieves 91.76%, 13.72% improvement than conventional work.

Keywords: qualitative action recognition, volleyball game analysis.

1. Introduction

Sports video analysis attracts lots of attentions due to the ballooning of sports video data. Among various kinds of

sports videos, volleyball game analysis becomes an attractive research target due to its representative condition, which consists of multiple players and complex background. Among various kinds of topics in volleyball analysis, player action recognition becomes essential since it serves as an elementary building brick for lots of applications, like player evaluation system and highlights extraction system. However most researches focus on the discrimination between different actions, i.e. to predict "which" action was performed at a specific time point. The quality of an action, which means how "well" the action was performed, has only received little attention so far, even though it potentially provides useful information for volleyball analysis. So, it is eager to develop a qualitative action recognition, which means to predict how well the action was performed. Our research target is to evaluate the volleyball action "receive" quality. Since normally "receive" is the beginning and end of one round game and influence the result most. So, it is more worth evaluating the "receive" than other actions. We judge the quality of "receive" from the return ball quality and motion quality. We classify the action "receive", into 4 categories based on the quality and recognize the input "receive" belonging to which quality category.

With the same target of play action recognition in sports, Guanyu Zhu⁽¹⁾ proposed an optical flow based motion representation and action recognition method for tennis player. But since for the qualitative action recognition, most actions share similar motion flows, so only based on optical flow is not robust for quantitative action recognition. With the same target of volleyball analysis, Hua-Tsung Chen⁽²⁾ proposed a basic action detection method based on the transitions of ball. However, most actions have similar ball trajectory transition. So, it is hard to judge the action quality only by ball trajectory.

Fig. 1. Framework

Kubota⁽³⁾ proposed a volleyball player action recognition method based on local motion flow. While due to the high similarity problem, different quality actions share the similar motion flows. So, this method is not robust enough for player qualitative action recognition in real volleyball game.

Most conventional work cannot work well due to the high similarity among different quality actions and appearance variation caused by the player change. For qualitative action recognition, we need to classify the action into different quality categories. It means the input actions belong to the same action category just with different qualities. They have similar global moving trajectory, body parts motion flow and so on. So, it is difficult to classify the actions which belong to the same category but with different qualities. The appearance variation means the appearance diversity in various target players. Since the physical data and personal playing habits of players are different from each other, this also causes difficulty in qualitative action recognition.

In this paper, we propose a ball motion state feature and abrupt pose feature based qualitative action recognition method for volleyball player. The quality of volleyball player action is affected by two factors: return ball quality and motion quality. From those 2 factors, we proposed 2 proposals. Ball motion state represents the ball motion state transformation during the action. This feature represents how the action acts on the ball and it can show the return ball quality and indicate the motion quality. Since this proposal extract features from the ball trajectory, so it eliminates the influence caused by the target player change. Abrupt player pose feature consists of 2 parts: abrupt hit

frame poses and abrupt pose variation. Abrupt hit frame pose feature get the player pose at the hit point. This feature represents the hit frame pose difference between standard action and nonstandard actions. Abrupt pose variation feature detects whether the player pose change or not during the action. Normally, a high-quality action has a standard and stable shape during the whole actions. It means there should not be abrupt pose variation during the whole actions. So, this feature detects whether there is an abrupt pose variation occurs or not during the actions. These parts are to classify action quality by the standard and stability. Finally, we combine those 2 parts to recognize the qualitative action.

The rest part of this paper is organized as follows: Section 2 presents the entire volleyball player qualitative action recognition method and details. Section 3 presents experiment results and analyses. The conclusion is drawn in section 4.

2. Proposal

2.1 Framework

The framework of qualitative action recognition is shown in Fig. 1. There are mainly two processes in this structure, one is ball motion state feature, the other is abrupt pose feature.

As for the ball motion state feature, a 3D ball tracking algorithm⁽⁴⁾ and 3D player tracking algorithm⁽⁵⁾ are employed to calculate the ball 3D trajectory and player 3D trajectory. Then we use the trajectories to detect the hit point. Next, the ball velocity before the hit point and after

the hit points are extracted based on least square fitting method. The ball motion state transformation represents how the action

Fig. 2. Ball motion state

acts on the ball and the velocity after the hit represents the final location of the ball. These features are used to represent the return ball quality and action efficiency.

As for the abrupt pose feature, it consists of 2 parts: hit frame abrupt pose and abrupt pose variation. Firstly, we detect the hit frame based on the distance between ball and player during the whole action. Then we use convex hull algorithm and geometric description to get the player pose feature. This proposal uses this feature to describe the player's pose. If an abrupt pose occurs, the action must be nonstandard. Secondly, we detect the player preparation pose and finish pose, then we calculate the pose variation during the whole action. If an abrupt variation happens, the action should be nonstandard. The abrupt frame pose represents the action standard and abrupt pose variation represents the stability of the action. By combining these two parts, the features contain the standard and stability of the action.

Finally, we employ a machine learning model, random forest⁽⁶⁾ to do training and prediction separately. Finally, we use weighted average to represent the final result. These two proposals will be introduced in section 2.2 and section 2.3 respectively.

2.2 Ball motion state feature

As the Fig. 2. shows, before the action, the ball comes from the rival side. After the action, the return ball destinations have lots of variation. It's obvious that some destinations are out of court, so this kind of action are bad quality. Some of them are going closed to the net, so this kind of actions are in good quality. Also, when a bad-quality action motion occurs, the ball motion state changed by the action is different from standard action. Since when the motion becomes nonstandard, it indicates some emergency occurs. Normally, such situation causes a

weird return ball motion state, including velocity, direction and trajectory. So, by extracting the ball motion state transformation, we can judge the return ball quality directly and evaluate the motion quality indirectly.

Table 1. Ball motion feature comparison

Action	X_{hit}^B	$V_{hit-before}$	$V_{hit-after}$
Return ball quality+	Center	Mid	Small
Return ball quality-	Boundary	Mid	Large
motion quality+	Center	Small	Mid
motion quality-	Boundary	Large	Large

+: good quality/-: bad quality

This proposal extracts ball motion feature around hit-point based on the ball trajectory. Features include the hit point position X_{hit}^B , the ball velocity before the hit point $V_{hit-before}$ and ball velocity after the hit $V_{hit-after}$. These features describe how the actions act on the ball and describe the ball motion after the action. For the hit point X_{hit}^B , normally the X_{hit}^B should be in the court or close to the player. But for the bad-quality action, sometimes the hit point X_{hit}^B is out the court or in the boundary of the court so that the player has to do actions nonstandardly like running or swoop to touch the ball. Also, for the $V_{hit-before}$, the same as X_{hit}^B , for bad-quality action, in most cases the $V_{hit-before}$ is rapid and tricky, but for good-quality the $V_{hit-before}$ is stable and slow. So, the hit point X_{hit}^B and $V_{hit-before}$ influence the motion quality somehow. For the $V_{hit-after}$, since after the hit, the final ball destination is only influenced by the gravity, so the difference of $V_{hit-after}$ represents the final location of return ball. The $V_{hit-after}$ helps to distinguish return ball quality is good or bad. Table 1 shows the feature' relationship.

The feature extracting processing is described here. Firstly, we define coordinate system as Fig .2 shows. The origin of coordinates is located at the center of the court. Define the 3D space as P and 3D position of ball at time k as X_k^B :

$$X_k^B = (x_k^B, y_k^B, z_k^B), (x_k^B, y_k^B, z_k^B) \in P, \quad (1)$$

also, we define the target player at time k as X_k^T :

$$X_k^T = (x_k^T, y_k^T, z_k^T), (x_k^T, y_k^T, z_k^T) \in P. \quad (2)$$

To get the hit point position X_{hit}^B , first we detect the hit frame. Since the ball and player get the closest at the hit frame, so the hit frame during the action $k \in \{1 \dots l\}$ can be calculated as:

$$\text{if } dist_i = \min_{i \in \{1..l\}}(dist(X_k^B, X_k^T)) \quad (3)$$

$$\text{hit frame} = i. \quad (4)$$

Fig. 3. Processing of hit frame pose extraction

For the $V_{hit-before}$ and $V_{hit-after}$, the difference represents how the ball motion state is transformed by the action. Through Eq. (3) (4), we get the hit point h during time $\{1 \cdots l\}$. Then, we use linear least square fitting method to approximate the 3D coordinate change according to time t .

Make X direction as an example:

$$v_{hit-before}^x = \frac{\sum t_i x_i - \sum t_i \sum x_i / k}{\sum t_i^2 - (\sum t_i)^2 / k}, t \in \{1 \cdots h\}, \quad (5)$$

$$v_{hit-after}^x = \frac{\sum t_i x_i - \sum t_i \sum x_i / k}{\sum t_i^2 - (\sum t_i)^2 / k}, t \in \{h \cdots l\}, \quad (6)$$

the other 2 directional velocity changes can be obtained in the same way. Finally, we get $V_{hit-before} = \{v_{hit-before}^x, v_{hit-before}^y, v_{hit-before}^z\}$ and $V_{hit-after} = \{v_{hit-after}^x, v_{hit-after}^y, v_{hit-after}^z\}$.

2.3 Abrupt pose feature

For different quality actions, the difference between high-quality action and low-quality action is the player pose. Normally, a high-quality action is performed with a standard

and stable shape. A bad-quality action is performed as a nonstandard pose and changed dramatically. So, from those 2 points: normal shape and stable shape, this paper proposes 2 features: abrupt hit frame pose and abrupt pose variation feature.

2.3.1 Abrupt hit frame pose feature

For the abrupt hit frame pose feature, the whole processing is like Fig .3. shows. First as equation (3) shows, we get the hit frame h during the whole action period. Then we pick up the image of hit frame in the video stream. After getting the picture, we do the dense feature sampling to get the feature points of target player, then using convex hull

algorithm to get the contour of player. In the Fig . 3, the feature points are drawn out. For the feature points, we using K-mean algorithm wo cluster them into 6 categories. Each category represent one part of the body. The different color for the points means different categories.

After we get the contour, we use 4 kinds geometric

Fig. 4. Comparison of different quality action

description to describe the contour of players, like fig .4. shows: 1) External ellipse (a, b, c), 2) Inscribed circle (o, r), 3) External rectangle (a, b, θ) 4) Convex polygon. The parameters of each geometric description are used as feature vector to do train and prediction.

Fig .4. shows the difference between standard motion shape and nonstandard shape. Firstly, the difference is the player contour. In standard motion, the player stands, and the lower body is like straddle. But for nonstandard, player swoop and stretch the body, so the external ellipse, the θ of external rectangle and other external information is different. The next difference is the inner body. For standard motion the player pose is standing but for nonstandard the player spread the body. So, for different quality actions, the inscribed circle is different.

For feature extracting processing, the dense points extraction part is as conventional work⁽³⁾⁽⁷⁾. Then this proposal uses convex hull algorithm to get the convex

polygon, based on the convex polygon we calculate kinds of geometric description. Since the location of player in the court will influence the numerical geometric data because the scale varies according to the distance between the camera and player, player location based feature normalization is conducted to eliminate the influence of the scale problem.

Fig. 5. Pose variation during the action

2.3.2 Abrupt pose variation feature

For the abrupt pose variation feature, like Fig .5. shows, for good-quality action, the player normally keeps the standard pose during the whole action. While in bad-quality action since in some situation, like player fall down or player jump to receive the ball like Fig .5., the ending pose of player becomes the very strange. So, this proposal is to extract the abrupt pose variation between the preparation pose and ending pose, use this feature to judge whether there is an obvious pose change during this action.

In this feature we pick up 5 feature variables as the feature vector: 1) convex polygon area variation ∇A , 2) convex length variation ∇L , 3) angel of external rectangle variation $\nabla \theta$ and 4) inscribed circle area variation ∇S , 5) the ratio of length side and short side of external rectangle σ .

Different features represent different variation. The convex polygon variation ∇A and ∇L shows the player contour scope change between the preparation pose and ending pose. The angle pf external rectangle angel variation $\nabla \theta$ shows the angle of player and floor change. When player swoops to the ball, the angel varies a lot. And the ratio of length side and short side of external rectangle σ shows the shape feature of contours. The inscribed circle area change ∇S represents the body inner shape changes or not.

Similar with hit frame pose feature, we do the dense feature points sampling as conventional wrok⁽³⁾⁽⁷⁾ to get the feature points. Then, this proposal uses convex hull

algorithm to get the contour and normalizes the value data based on the player position. Finally, this proposal calculates the pose geometric feature separately. A', θ', L, S' represent the geometric description of preparation pose and A'', θ'', L'', S'' represent the ending pose. Eq. (7), (8), (9). (10) show how to calculate the pose variation. The ratio of length side and short side of external rectangle $\sigma = \frac{a}{b}$.

$$\nabla A = A' - A'', \quad (7)$$

$$\nabla \theta = \theta' - \theta'', \quad (8)$$

$$\nabla L = L' - L'', \quad (9)$$

$$\nabla S = S' - S''. \quad (10)$$

3. Experimental result

3.1 Experimental condition and evaluation criteria

The experiment is based on multi-view videos of Semifinal Game and Final Game of 2014 Japan Inter High School Games of Men's Volleyball in Tokyo Metropolitan Gymnasium. The cameras are set at each corner of the court. Video resolution is 1920*1080, frame rate is 60 frames per second, and shutter speed of the camera is 1000 frames per second. The latter parameter prevents motion blur in video sequence. The experiments are conducted by programming with C++ language and OpenCV 2.4.10 on a machine of 3.60GHz CPU and 8GB RAM.

In this research, three classical evaluation criteria defined in reference⁽⁸⁾, accuracy, recall and precision are used to evaluate the results. Accuracy reveals how many samples are correctly predicted. Recall is used to measure the ability of this detection system for extracting relevant information. Precision is used to describe the irrelevant information rejection ability of this system.

3.2 Datasets and Qualitative actions

The semifinal game and final game are named as Game A and Game B, we classify the “receive” into 4 categories based on 2 factors: return ball quality and motion quality. We define 4 qualitative action categories: Quality++, Quality+, Quality-, Quality--. For the return ball quality, as Fig. 6. shows, a definition of efficient region of “receive” return ball is given by the conventional work⁽⁹⁾. Based on the definition, we define the good and bad quality of return ball. If the return ball reaches the A or B region, we regard it as good quality and others are bad quality. For the action motion quality, in Fig .6. two models are defined:

standard motion and nonstandard motion. If a “receive” is performed as standard motion, we define it good quality, otherwise bad quality. By combining those two factors, we define 4 qualitative action categories. The whole dataset is like table 2 Shows.

3.3 Result and Discussion

The experiments are conducted in 2 models, Training Game_A, testing Game_B and training Game_B, testing Game_A. In the table 3, 4, the average result experiment

Table 2. Datasets of qualitative actions

Action	++	+	-	--	Total
Game_A	45	18	20	5	88
Game_B	57	19	9	6	91

++: Standard motion/Good return ball;

+: Nonstandard motion/Good return ball;

-: Standard motion/Bad return;

--: Nonstandard motion/Bad return;

Fig. 6. Action motion quality

results are described.

The conventional work is based on a 2D flow based volleyball player action recognition work (3). Since the conventional work does not focus on the quality evaluation, it does not consider the return ball quality and motion

standard and stability. So, the performance of conventional work on the quality evaluation is not good, especially the recall. It means the conventional work cannot extracting relevant information of bad return ball quality and bad motion quality. It cannot classify the qualitative actions which are extremely similar with each other.

Thanks to the player pose variation, the inner shape and outer contour difference is more obvious rather than the motion flow feature. Since the “receive” shares the same body motion flow of arm or foot, so only the motion flow is not enough to distinguish different qualitative “receive”. But

Table 3. Learn Game_A and Test Game_B

Action	Criteria	conventional	P2	P1+P2
Quality++	accuracy	75.28%	68.54%	93.18%
	recall	77.78%	100%	95.55%
	precision	74.47%	61.64%	87.70%
Quality+	accuracy	81.11%	78.65%	94.31%
	recall	11.11%	27.78%	72.22%
	precision	66.67%	45.45%	72.22%
Quality-	accuracy	76.40%	83.15%	92.05%
	recall	20.00%	30.00%	65.00%
	precision	44.44%	85.71%	61.90%
Quality--	accuracy	80.48%	78.88%	90.00%
	recall	0%	20.00%	60.00%
	precision	0%	25.00%	60.00%

according to the game situation in real game, the nonstandard actions perform like spreading the body or curling the body. So, although the flow of arm, like waving flow, is similar to each other but the pose is different. Also, an abrupt pose variation happens in a bad-quality action. In a bad-quality action, the pose variation during the action is much more obvious than a good-quality action. The obvious feature helps us to distinguish the bad-quality motion and good-quality motion.

The 3D ball motion feature also helps to improve the final performance. As we analysis before, the $V_{hit-after}$ feature indicates the final location of the return ball, so the difference between the feature distinguishes the good return ball quality and bad return ball quality. Also, according to the volleyball game situation analysis, the ball hit point and $V_{hit-begin}$ have tight relationship with the motion quality. The hit point and $V_{hit-begin}$ indicate the abrupt level of the action. The more abrupt the receive is, the motion quality is tended to be bad.

By combining those features, ball motion feature and

abrupt pose feature promote each other mutually and good-quality receives and bad-quality receives are distinguished well.

4. Conclusions

In this paper, we propose a ball motion state and player pose based player qualitative action recognition method. Through the ball motion state, we evaluate the action the return ball quality and the motion quality. Player pose feature is to extract the hit frame pose and pose variation during the whole action. These two features represent the standard and

Table 4. Learn Game_B and Test Game_A

Action	Criteria	conventional	P2	P1+P2
Quality++	accuracy	57.14%	75.90%	89.47%
	recall	45.28%	92.31%	96.49%
	precision	77.42%	85.34%	90.16%
Quality+	accuracy	76.19%	81.92%	93.40%
	recall	21.43%	50.00%	73.68%
	precision	25.00%	46.67%	70.00%
Quality-	accuracy	78.89%	89.28%	96.70%
	recall	35.00%	44.44%	66.67%
	precision	100.00%	50.00%	66.67%
Quality--	accuracy	82.18%	83.89%	85.00%
	recall	16.67%	33.33%	66.67%
	precision	16.67%	40.00%	50.00%

stability of action. The experiment result shows the average accuracy of each qualitative action categories achieve 91.76%, 13.72% improvement than 2D flow based action recognition method.

In the future, we hope to apply our method to other action in volley ball game like: spike, toss, serve and block, which is also useful for volleyball game analysis. Furthermore, hardware based acceleration is also an important research target.

Acknowledgment

This work was supported by KAKENHI (16K13006) and Waseda University Grant for Special Research Projects (2018K-302).

References

(1) Guangyu Zhu, Changsheng Xu, Qingming Huang, Wen

Gao, "Action Recognition in broadcast Tennis Video", IEEE 18th International Conference on Pattern Recognition, pp. 251-254, Aug, 2006.

- (2) Hua-Tsung Chen, Hsuan-Shen Chen and Suh-Yin Lee, "Physics-based Ball Tracking in Volleyball Videos with its Applications to Set Type Recognition and Action Detection", IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1097-1100, April 2007.
- (3) E. Kubota, M. Honda and T. Ikenaga, "Action detection of volleyball using features based on clustering of body trajectories", Technical report of IEEEJ, vol. 272 (in Japanese) iii, 35, 36, Feb
- (4) X.N Cheng, N. Ikoma, M. Honda and T. Ikenaga, "Anti-occlusion observation model and automatic recovery for multi-view ball tracking in sports analysis", 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016), Mar 2016
- (5) S.Y Huang, X.Z Zhuang, N. Ikoma, M. Honda and T. Ikenaga, "Particle Filter with Least Square Fitting Prediction and Spatial Relationship based Multi-view Elimination for 3D Volleyball Players Tracking", 12th IEEE Colloquium on Signal Processing & its Applications (CSPA), pp 28-31, Mar 2016.
- (6) A. Liaw and M. Wiener, "Classification and Regression by RandomForest", R news, vol.2, pp.18-22, Dec 2002.14, 44.
- (7) H. Wang, A. Kläser, C. Schmid and C.L. Liu, "Action Recognition by Dense Trajectories," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp.3169-3176, Jun 2011.iii, 7, 14, 20, 33, 34.
- (8) T. Fawcett, "An introduction to ROC analysis", Pattern Recongn. Lett, vol.27 no.8, pp. 861-874, 2006.
- (9) Castro J. Mesquita I.(2010) Analysis of the attack tempo determinants in volleyball's complex II - a study on elite male teams. International Journal of Performance Analysis in Sport 10(3), 197-206