Start and End Point Detection of Weightlifting Motion using CHLAC and MRA

Fumito Yoshikawa[1], Takumi Kobayashi[1], Kenji Watanabe[1]
Katsuyoshi Shirai[2] and Nobuyuki Otsu[1]

[1] AIST Tsukuba, C2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

[2] Japan Institute of Sports Sciences, 3-15-1 Nishigaoka, Kita-ku, Tokyo 115-0056, Japan

Abstract. Extracting human motion segments of interest in image sequences is essential for quantitative analysis and effective video browsing, although it requires laborious human efforts. In analysis of sport motion such as weightlifting, it is required to detect the start and end of each weightlifting motion in an automated manner and hopefully even for different camera angle-views. This paper describes a weightlifting motion detection method employing cubic higher-order local auto-correlation (CHLAC) and multiple regression analysis (MRA). This method extracts spatio-temporal motion features and leans the relationship between the features and specific motion, without prior knowledge about objects. To demonstrate the effectiveness of our method, the experiment was conducted on data captured from eight different viewpoints in practical situations. The detection rates for the start and end motions were more than 94% for 140 data in total even for different angle views, 100% for some angles.

1 Introduction

Detecting and segmenting human action and behavior of interest in video sequences is necessary in various applications such as quantitative motion analysis and video browsing. It is a fundamental procedure for understanding the motions in question.

In sport motion analysis, especially for weightlifting, athlete's sport-motions are often analyzed for improving their sport performance by using video sequences captured during competition and training. In biomechanical studies of weightlifting, much research efforts have been made to explore the relations between several kinematic parameters, such as time from barbell lift-off to its maximum height, and the winning lifts by conventional approaches involving manual indexing operations [1, 2]. In practical situations, however, quick feedback of the resultant quantitative data and/or videos is expected to be conducted for the relevant coaches and athletes without using any intrusive manner in data acquisition, for example by placing markers on the human body. In order to reduce the burden imposed on human operators for more detailed analysis, automation of motion analysis is required. Thus, first of all, automatic detection of the start and the end of predefined single motions

Yoshikawa F., Kobayashi T., Watanabe K., Shirai K. and Otsu N. (2010).
Start and End Point Detection of Weightlifting Motion using CHLAC and MRA.
In *Proceedings of the 1st International Workshop on Bio-inspired Human-Machine Interfaces and Healthcare Applications*, pages 44-50
DOI: 10.5220/0002813100440050

(the snatch, or the clean and jerk) in weightlifting is essential for both the quantitative analysis and the effective video handling.

For the task to detect and recognize the full-body human motions, many researchers have investigated the performance of various video-based motion analysis methods [3 - 6]. These methods require segmentation of the target objects such as persons, in which the segmentation error tends to affect final recognition. In addition, the conventional approaches include sequential and procedural processes require too special and tedious steps. These make it difficult to design adaptive and real-time systems.

In recent years, on the other hand, a scheme of adaptive vision system has been presented, which comprises two stages of feature extraction, namely, higher-order local auto-correlation (HLAC) or its extension CHLAC (Cubic HLAC) [7] and multivariate analysis [8, 9]. Concerning human motion and behaviour analysis, CHLAC approach has been successfully applied to motion recognition [7], unusual motion detection [10, 11] and motion segmentation [12]. The CHLAC approach, however, has not been applied to detection of predefined motions. In [12], the task of segmenting single weightlifting motions into detailed primitive motions is addressed in the experiment; however, the segmentation methods proposed there need image sequences to be clipped so as to include the entire weightlifting action of interest in advance.

In this paper, we applied CHLAC and MRA to the start and end time point detection in weightlifting, obeying the simple statistical scheme framework [9]. In the experiment we used the dataset which had been acquired by capturing national elite athletes' lifts from eight different viewpoints in practical situations. The experimental results demonstrated the effectiveness of the present method.

2 Method of Weightlifting Motion Detection

The proposed method consists primarily of three steps; preprocessing, motion feature extraction, and linear regression. In this section, we begin by making brief explanation of the input images and the tasks to be addressed.

2.1 Input Image and Task

Weightlifting videos captured during competition and training usually contain both of transient barbell-lifting segments ("work") in question and the others ("rest") including setup of barbell weight. The "work" and "rest" segments are alternatively concatenated. In addition, the videos are often captured from different viewpoints in practical environments. Note that they also contain background noise derived from other moving persons. Fig. 1 shows examples of actual still images in weightlifting videos captured from different viewpoints during training.

In order to clarify the task addressed in this paper, examples of multi-viewed image sequences around the start and the end motion in weightlifting are illustrated in Fig.1. These motion segments are defined in this study as follows; the start motion is from the time when the barbell plate lifts off the floor until the bar reaches maximum height above the platform surface while the lifter is moving into squat position to catch it, on the other hand, the end motion is from the time when the barbell starts to descend until it once reaches the floor. In practical applications for in-depth motion analysis, detecting and indexing the time just when the barbell plates are lifted off is the primary procedure. The rest of this section describes the proposed approach to automatic detection of these start and end motions in weightlifting.

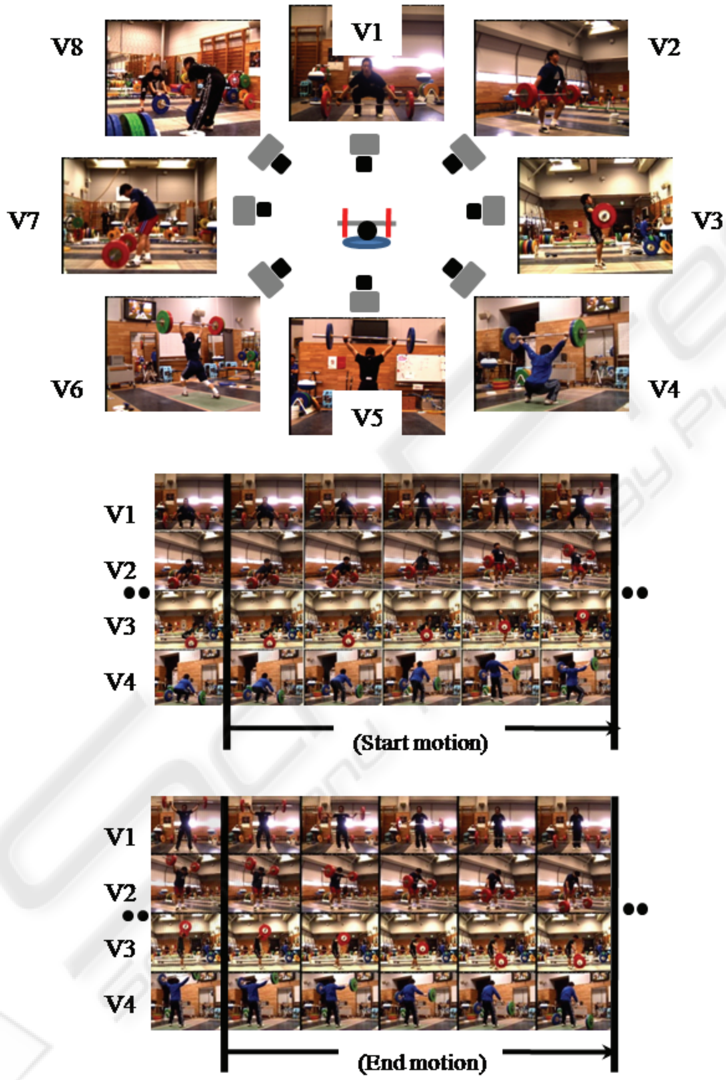

Fig. 1. Examples of video captured during weightlifting training: eight angle-views (*upper*), image sequences around the start (*middle*) and the end (*bottom*) of lifting motions.

2.2 Preprocessing

In the preprocessing, we apply frame differencing and then automatic thresholding [13] in order to detect and binarize motion pixels as in [7]. These processes filter out both inherent noise and brightness information, which are irrelevant to the motion itself. Consequently, pixel values in each frame become either 1 (moved) or 0 (static). The examples of the binary images are shown in Fig.2 at the same scenes as the middle and bottom of Fig.1.

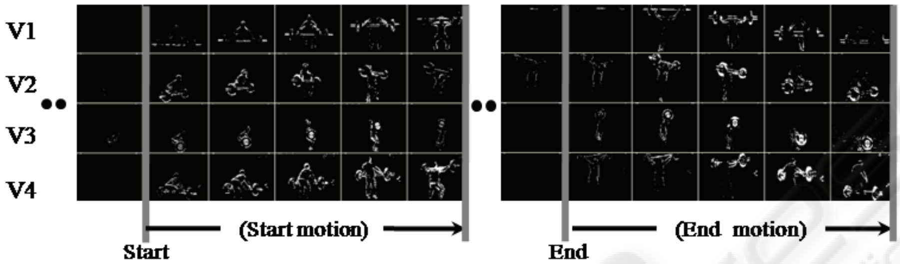

Fig. 2. Examples of the preprocessed image sequences around the start and the end of lifting motions.

2.3 Motion Feature Extraction

In the stage of feature extraction, we employ Cubic Higher-order Local Auto-Correlation (CHLAC) [7]. CHLAC enables simultaneous extraction of spatio-temporal features from the motion image. Let $f(\boldsymbol{r})$ be three way data defined on the region (cubic data) $D : X \times Y \times 3$ with $\boldsymbol{r}= (x, y, t)^T$, where X and Y are the width and height of image frame and T is the length of a time-window. Then, the N-th order auto-correlation can be defined as,

$$x_N(t;\boldsymbol{a}_1,\boldsymbol{a}_2 \cdots ;\boldsymbol{a}_N)=\int_T\int_{X,Y} f(\boldsymbol{r})f(\boldsymbol{r}+\boldsymbol{a}_1)f(\boldsymbol{r}+\boldsymbol{a}_2)\cdots f(\boldsymbol{r}+\boldsymbol{a}_N)d\boldsymbol{r} \tag{1}$$

where the $\boldsymbol{a}_i$ ($i = 1, \cdots, N$) are displacement vectors from a reference point $\boldsymbol{r}$. Since Eq. (1) can take many different forms by varying N and $\boldsymbol{a}_i$, we limit $N \leq 2$ and $\boldsymbol{a}_i$ to a local region: the configurations of $\boldsymbol{r}$ and $\boldsymbol{a}_i$ are represented as mask patterns shown in Fig.3. The motion features are extracted by scanning the entire data set D with local cubic mask patterns. Thus, CHLAC feature corresponds to a histogram of local configuration patterns (auto-correlation) of moving points (pixels) found by frame difference. The dimension of CHLAC of up to the second order within the local 3x3x3 region is 251 for the binary data. CHLAC has a parameter denoted by Δr which is the spatial interval of the mask patterns along the x- and y- axes in the image frame.

CHLAC features possess important properties of shift invariance (rendering the method segmentation-free) and robustness to noise in data. Moreover, this method requires no prior knowledge or heuristics about objects. These favourable properties can benefit all aspects of approach to adaptively detect weightlifting motions including possible variability in terms of their appearances due to difference in lifter's physical attribute, kinematic profile and camera angle.

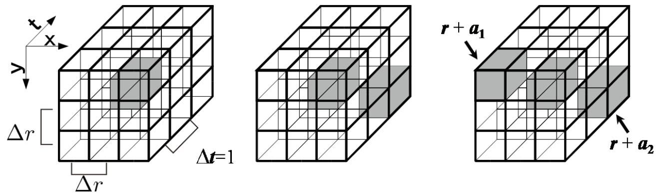

Fig. 3. Examples of mask patterns: (left) N=0; (middle), $N = 1$, $a_1 = (\Delta r, \Delta r, 1)^T$; (right), $N = 2$, $a_1 = (-\Delta r, -\Delta r, -1)^T$, $a_2 = (\Delta r, \Delta r, 1)^T$.

2.4 Linear Regression

In the training phase, effective features for the start and end motion detection are extracted from the given training example. The pairs of the motion feature vector x_i and the teacher signal c_i at time i are given. We apply multiple regression analysis (MRA), which determines the optimal linear coefficients $\boldsymbol{a}$, to estimate c from x: $c \approx a'x + b = \hat{a}'\hat{x}$, where b is constant, $\hat{a} = [a', b]'$, $\hat{x} = [x', 1]'$. In this study, the teacher signals are binary, assigning 1 at times during the start and the end motions, and otherwise assigning 0.

Given the motion feature $\boldsymbol{x}$, the existence of the target motion segment can be estimated by $c = a'x + b$. In the method, the target motions are finally identified by detecting the local peak along the time axis and thresholding it after applying moving average to the estimated values over a time-window T.

3 Experiments

The proposed method was applied to automatic detection of weightlifting motions from the image sequences. The dataset utilized in this experiment comprises 140 video sequences of the successful *snatch* and *clean and jerk* performed by national top-level athletes in different categories according to their bodyweight. The dataset had been acquired by filming lifts from eight different viewpoints in practical situations, as shown in Fig. 1. These data were captured at 30 frames per second (fps) and 320 x 240 pixels (QVGA).

For evaluating the performance of the proposed method, a leave-one-out scheme was applied to video sequences captured from the same camera angle, respectively, and then precision rates were calculated for both each target motion, i.e. start and end motion and each camera angle. In this evaluation, the detected point was regarded as correct if it was within each time duration of the target motion which was strictly determined by hand as ground truth.

CHLAC features are obtained by using all mask patterns of $\Delta r = 1,3,5,7,9$. The time-window T for smoothing the estimation results is 27 and 16 for the start and end detection, respectively, based on the averaged time interval of the target motions in the dataset. The results are shown in Table 1. The proposed method produces favorable results on every angle views and different kinds of weightlifting, the *snatch* and the *clean and jerk*. These results can indicate that our method can address the corresponding needs in the practical situations of weightlifting. The degradation of the precision rate for the end motion detection in V1 was largely due to that the vertically higher-positioned barbell was out of frame-view in several input images. In addition, some sample in the dataset includes an incompletes single lifting movement because the start or end of a single weightlifting motion is just near the corresponding start or end of a clipped video sequence.

Table 1. Detection rates (%) over different angle-views for each start and end motion in weightlifting. Angle views in this experiment nearly correspond to those illustrated in Fig. 1.

Angle-View	V1	V2	V3	V4	V5	V6	V7	V8
Start	100	100	94.7	94.1	100	94.4	100	94.7
End	71.4	100	89.5	100	94	88.9	90.0	93.8

From the viewpoint of kinematics, the entire single lift can be subdivided into several phases and the profiles in each phase are different among athletes, as indicated in [1]. In order to cope with the diversity in the kinematic profiles, our method employs various sizes of mask patterns for motion feature extraction, which can contribute to the performance. On the other hand, the motion orientation in the pulling phase after the start of each lift is similar to that after squat position to catch the barbell and that during jerk thrust, and consequently the spatio-temporal features of these movements extracted locally along the time axis can be not largely varied in some cases.

We applied this method to other sport motion, such as service detection of badminton, and obtained the similar results, which shows the validity and generality of our method [14].

4 Concluding Remarks

We have presented CHLAC approach to automatic detection of weightlifting motions, which can be mentioned as typical examples of predefined transient motions. The present method consists of motion feature extraction by CHLAC and prediction by MRA, and yields favorable detection performances for the start and end motions in weightlifting. By detecting these two motions, the whole weightlifting motion can be roughly segmented. Then, the weightlifting motion can be finely segmented, such as by using the methods proposed in [12]. It is expected that the integration between these approaches can contribute to more precise analysis of single transient motions of interest which are not limited to weightlifting motions.

Acknowledgements. We thank Mr. Miyoji Kikuta and Mrs. Kumiko Haseba at Japan Weightlifting Association for his cooperation on data gathering.

References

1. Garhammer, J.: Biomechanical Profiles of Olympic Weightlifters. International Journal of Sport Biomechanics, 1, (1985), 122-130
2. Schilling, B. K., Stone, M. H., O'bryant, H. S., Fry, A. C., Coglianese, R. H., Pierce, K. C.: Snatch technique of collegiate national level weightlifters. Journal of Strength and Conditioning Research, 16 (4), (2002), 551–555
3. Bobick, A. F., Davis, J. W.: The Recognition of Human Movement Using Temporal Templates. In: IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 23, No. 3, (2001) , 257-267
4. Zelnik-Manor, L., Irani, M.: Statistical Analysis of Dynamic Actions. In: IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 28, No. 9, (2006), 1530-1535
5. Zhu, G., Xu, C, Huang, Q, Gao, W, Xing L.: Player Action Recognition in Broadcast Tennis Video with Applications to Semantic Analysis of Sports Game. In: the 14th annual ACM international conference on Multimedia, (2006)
6. Roh, M.C., Christmas, B., Kittler, J., Lee, S.W.: Gesture Spotting for Low-resolution Sports Video Annotation. Pattern Recognition 41, (2008), 1124-1137
7. Kobayashi, T., Otsu, N.: Three-way Auto Correlation Approach to Motion. Recognition Pattern Recognition Letter, Vol. 30, No. 3, (2009), 212- 221
8. Otsu, N., Kurita, T.: A New Scheme for Practical Flexible and Intelligent Vision System. In: IAPR Workshop on Computer Vision, (1988)
9. Otsu, N.: CHLAC Approach to Flexible and Intelligent Vision Systems. In: ECSIS Symposium on Bio-inspired, Learning, and Intelligent Systems for Security, (2008)
10. Nanri, T., Otsu N.: Unsupervised Abnormality Detection in Video Surveillance. In: IAPR Conference on Machine Vision Applications, (2005)
11. Iwata, K., Satoh, Y., Sakaue K., Kobayashi, T., Otsu, N.: Development of Software for Real-time Unusual Motions Detection by Using CHLAC. In: ECSIS Symposium on Bio-inspired, Learning, and Intelligent Systems for Security, (2008)
12. Kobayashi, T., Yoshikawa, F., Otsu, N.: Motion Image Segmentation Using Global Criteria and DP. In: International Conference on Automatic Face and Gesture Recognition, (2008)
13. Otsu, N.: A threshold selection method from gray-level histogram, In: IEEE Transactions on System Man Cybernetics, Vol. SMC-9, No. 1, (1979) , 62-66
14. Yoshikawa, F., Kobayashi, T., Watanabe, K., Otsu, N.: Automated Serve Scene Detection for Badminton Game Analysis Using CHLAC and MRA. In: Asian Conference on Physical Education and Computer Science in Sports, (2009)