Fast Online Action Recognition with Boosted Combinational Motion Features

Masamichi Shimosaka, Takayuki Nishimura, Yu Nejigane, Taketoshi Mori, and Tomomasa Sato
Graduate School of Information Science and Technology
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
{*simosaka, nisimura, tmori, nejigane, tomo*}*@ics.t.u-tokyo.ac.jp*

***Abstract*— In this paper, we propose a fast and robust online action recognition method. The main features of the proposed method are: 1) to select a small number of critical motion features from a very large set of motion feature templates and to release humans from task of designing critical motion features, 2) to require very small calculation cost for recognition compared to conventional methods, 3) to exploit "Combinational Motion Features" which we propose as a new conception so as to construct a robust action recognizer.**

We evaluated the proposed method to gait action recognition, such as *walking* and *running*, by utilizing motion capture data. In the result, the proposed method reduced parameters given by human to action recognizer and lessened human's task. In addition, the proposed method needed very small calculation cost for recognition, and can recognize robustly as much as conventional action recognition method based on support vector machine. Moreover, the introduction of combinational motion features enhanced recognition performance.

I. Introduction

It is significant for realization of supporting humans by robots to recognize humans actions [1], [2], [3]. While, it is necessary for the achievement of immediately support that robots can obtain human motion data online and analysis promptly it. Accordingly, we proposed the method for online daily life action recognition so as to construct online motion recognition system.

There are a number of researches on action recognition [4], [5]. For example of recognition based on hidden Markov model (HMM), Inamura et al. [6] proposed the scheme for recognizing sequential human motion of his or her joint angle sectioned with uniform interval. While, for example of recognition using support vector machine (SVM), Cao et al. [7] extracted features from video clips recorded in a direction perpendicular to person's trajectory and recognized 8 types of gait motions. In our previous work [8], we constructed the recognition system for daily life actions, whose input is data fetched by motion capture system.

These methods with machine learning techniques such as HMM and SVM have high recognition performance, but generally need many parameters to enhance recognition performance. And these parameters must be given by humans in advance. In addition, calculation cost for recognition is large, and then it is difficult to realize online motion recognition. Moreover, the methods make humans design important features for action recognition despite the designing features is difficult and bothers humans.

In order to solve these problems, we adopt boosting approach; an ensemble learning algorithm. Ensemble learning algorithm [9] is to generate various and simple learning models based on different training samples. Then the algorithm constructs recognizer by joining together simple learning models. The constructed recognizer is robust comparable to a complicated and robust recognizer.

Especially in ensemble learning algorithm, boosting [10] is generally more robust than other ensemble learning algorithm, such as bagging [11]. Boosting generates simple classifiers called weak learners in stages which trained with the data. Each of weak learners is simple and low performance for recognition. However, boosting algorithm construct a robust classifier by joining together weak learners.

In classification based on boosting, one of the most important issues is how to design weak learners. For example, in the domain of pedestrian detection, Viola et al. [12] designed the weak learners classifying by threshold processing. Target for threshold processing is the difference between the sums of the pixels within two regions and successive two images. Calculation cost is very small in each weak learner because the weak learner is simple. Viola et al. designed such weak learners in order to construct a robust classifier while retaining framework of small calculation cost.

In this paper, online action recognition is constructed with model based method. We are inspired by the above works of boosting. And then we design simple weak learners, which is called "Action Cue". An action cue classifies by threshold processing of "Fundamental Motion Features". Fundamental motion features are calculated from measured motion data such as joint angle and position by simple forward kinematics. Thereby each stage of the boosting process that selects a new action cue can be viewed as a feature selection process. That is to say, a few important features for recognition are automatically selected in boosting process. In addition, each action cue has very small calculation cost. Because action cues extract motion features by forward kinematics and classify by threshold processing.

However, there are actions which are not classified robustly by a classifier with only fundamental motion features. Thus, we introduce "Combinational Motion Feature" which is the

combination of several fundamental motion features so as to classify robustly actions which are not easy to be classified with only fundamental motion features. In detail we describe in section Ⅲ.

Based on above discussion, we propose a fast and robust online action recognition method in this paper. The proposed method has the following three features; 1) boosting process selects automatically a small number of important motion features from a huge library of motion feature templates and release humans from task of designing important motion features, 2) the proposed method needs very small calculation cost for recognition because each action cue is simple, 3) the introduction of "Combinational Motion Feature", which is the combination of several fundamental motion features, allows the actions, which are not easy to recognize with only fundamental motion features, to be robustly recognized.

This paper is organized as follows. Section Ⅱ discusses action recognition based on boosting algorithm. Section Ⅲ details designing action cues. Section Ⅳ describes some experimental results, including a detailed description of our experimental methodology. Finally section Ⅴ contains a discussion of the proposed method and future works.

Ⅱ. Action Recognition Based on Boosting

A. Scheme of online action recognition

Daily life actions have a different characteristic from other actions, such as gesture and sign language. Actions of gesture and sign language are exclusive in relationship among these actions, that is, such actions never occur simultaneously. However, daily life actions are not always exclusive in relationship among these actions, in turn, several actions may occur simultaneously. For example, humans can recognize the two actions involved when observing someone is *looking up* and *walking*.

Therefore we construct the action recognition system that can simultaneously label motion data with several actions. Fig 1 shows the processing flow and the structure of our recognition system. Sequential measured motion data is utilized as input of the proposed recognition system. In order to realize the simultaneous recognition, the system contains multiple classifiers, each of which is assigned to classify one specific action. The process of each classifier runs in parallel with and independent of the others. The recognition performance of system depends largely on the recognition performance of each action classifier, and then modeling action classifiers is very important.

B. Modeling action binary classifiers with boosted

In this paper, we simply design an action classifier to classify by threshold processing of a scalar motion feature in order to construct a robust action classifier which needs small calculation cost for recognition.

That is to say, if a scalar value of motion feature is no less than or no more than given threshold, a classifier recognize that action assigned the classifier occurs or not.

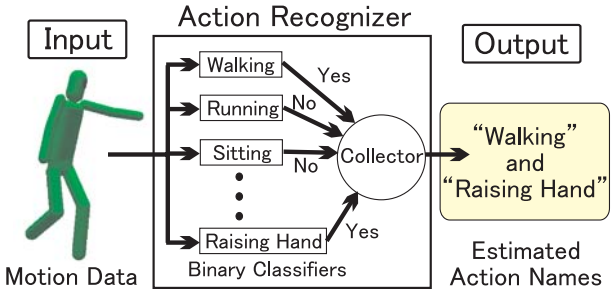

Fig. 1. Scheme of Online Motion Recognition

Another reason we construct such classifiers is that a scalar value of motion feature can represent certain relationship among body parts. For example in case of recognition of *lying*, the height of head for hip represent a characteristic relationship among body parts in *lying*. And the classifier classifies human to be *lying* if the height is not more than the given threshold. We call such a classifier "Action Cue". Fig 2 shows the processing flow of action cue in *lying* classification. Action cues will be detailed in section Ⅲ.

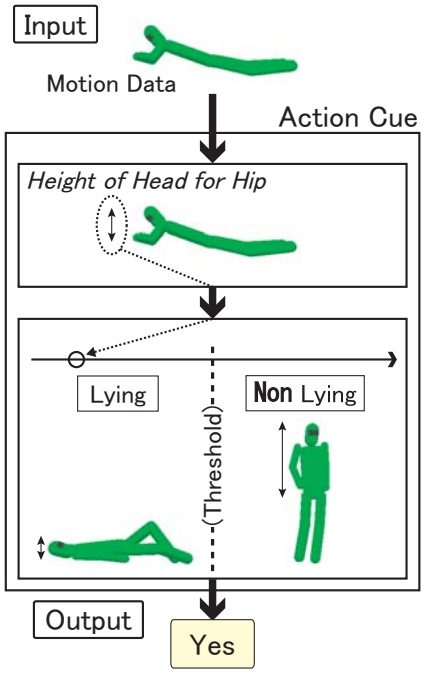

Fig. 2. Example of Processing Flow of Action Cue

However, an action cue can hardly classify robustly. Therefore, we joint low performance action cues in order to construct a robust action classifier while retaining small calculation cost. These jointed classifiers classify by weighted majority vote based on recognition confidence of each action cue.

The robust classifier $H(\boldsymbol{x})$ is described as follows:

$$H(\boldsymbol{x}) = \begin{cases} +1, & \sum_{k=1}^{K} \alpha_k h_k(\boldsymbol{x}) \geq 0 \\ -1, & \text{otherwise} \end{cases} ,$$

where, $H(\boldsymbol{x}) = +1$ or -1 denote assigned action occur or not at the moment, and $\boldsymbol{x}$ is measured motion data such as joint angle and position. While, for $k = 1, \ldots, K$, h_k is k-th action cue and α_k is recognition confidence of k-th action cue. Fig 3 shows the processing flow and the structure of the action classifier $H(\boldsymbol{x})$.

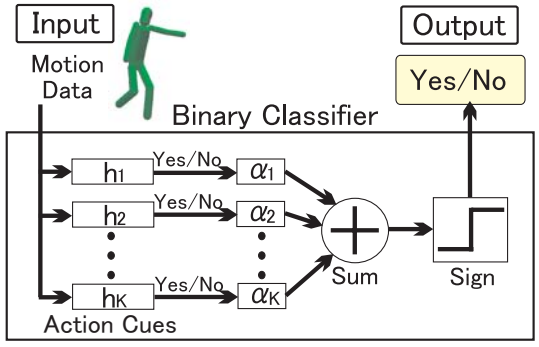

Fig. 3. Configuration of Action Classifier

In this paper, we decide α_k by adopting AdaBoost; a boosting algorithm. AdaBoost is widely used and typical algorithm in boosting algorithm. Table I shows our used AdaBoost algorithm. In the rest of this paper, the term "boosting" represents the AdaBoost learning algorithm.

III. Designing action cues for recognition

This section describes how to design action cues and then introduces combinational motion features (CMFs); powerful tools for action recognition in the proposed method.

A. Motion features utilized by action cues

As noted in section I, online action recognition system is constructed with model based method in this paper. Thus the action recognition system momentarily measures joint angle and position by leveraging image processing or information of sensors, such as attitude sensors and acceleration sensors. In this paper, we adopt a magnetic motion capturing system to fetch motion data.

Then action cues classify by using instantaneous motion features which are calculated from measured motion data by simple forward kinematics. However, the system doesn't use motion features that are extracted with considering time-series of motion data. For examples, in gait motion recognition, one of these motion features is frequency transformation of stride. Frequency transformation of stride as a motion feature is calculated by frequency transformation from set of distances between right foot and left foot per unit of measure within a constant long time interval. We utilize only instantaneous motion features and never utilize motion features considering time-series of motion data, since the latter motion features need long time delay to be extracted and then the delay thwarts realizing fast online action recognition. In addition, since humans can almost recognize actions by just looking a picture clipping human actions and without considering time-series of motion data, we expect that recognition system also can recognize by utilizing only instantaneous motion features.

TABLE I

The AdaBoost Algorithm for Modeling Each Action Classifier

0 Given as training data:
$(\boldsymbol{x}^{(1)}, y^{(1)}), \ldots, (\boldsymbol{x}^{(n)}, y^{(n)}), \ldots, (\boldsymbol{x}^{(N)}, y^{(N)})$;
$\boldsymbol{x}^{(n)} \in X$, $y^{(n)} \in \{+1, -1\}$
if action occur, $y^{(n)} = +1$ otherwise $y^{(n)} = -1$

1 Initialize

$$D_1(n) = \frac{\exp(y^{(n)} \log \sqrt{\frac{N_\mathrm{n}}{N_\mathrm{p}}})}{\sum_{n=1}^{N} \exp(y^{(n)} \log \sqrt{\frac{N_\mathrm{n}}{N_\mathrm{p}}})},$$

where N_p is a # of data with $y^{(n)} = +1$, N_n is a # of data with $y^{(n)} = -1$, and $D_k(n)$ is weight of n-th data at k-th round.

2 For $k = 1, \ldots, K$:

- Select action cue h_k which minimizes the error rate

$$\epsilon_k = \sum_{n:h_k(\boldsymbol{x}^{(n)}) \neq y^{(n)}}^{N} D_k(n)$$

- Update the weights:

$$D_{k+1}(n) = \frac{D_k(n) \exp(-\alpha_k y^{(n)} h_k(\boldsymbol{x}^{(n)}))}{Z_k},$$

where Z_k is a normalization factor.

3 Output the classifier:

$$H(\boldsymbol{x}) = \mathrm{sgn}\Big(\sum_{k=1}^{K} \alpha_k h_k(\boldsymbol{x})\Big),$$

where $\alpha_k = \frac{1}{2} \log\Big(\frac{1-\epsilon_k}{\epsilon_k}\Big)$.

B. Basic policy for designing action cues

In order to lessen calculation cost for recognition, we should design action cues to be simple. Therefore, we utilize the action cues used in our previous work [13]. The action cues classify by threshold processing of a scalar motion feature. The action cue is described as follows:

$$h_k(\boldsymbol{x}) = \begin{cases} +1, & [\![\beta_k \phi_{\mathrm{fmf},k}(\boldsymbol{x}) \geq b_k]\!] = 1 \\ -1, & \text{otherwise} \end{cases},$$

where $\boldsymbol{x}$ is measured motion data as input of action cues, and b_k is threshold of k-th action cue. While $\boldsymbol{x}$ is vector of 30 to 60 dimensions $\boldsymbol{x} = \{\boldsymbol{P}, \boldsymbol{\Theta}\}$, where $\boldsymbol{P}$ is position x,y,z of root body part in world coordinate system, $\boldsymbol{\Theta}$ is set of joint angle θ for each body part (Fig 4). $\phi_{\mathrm{fmf},k}$ is the function which extract a scalar fundamental motion feature (FMF) $r_k \in \mathbb{R}$ from measured motion data $\boldsymbol{x}$ by simple forward kinematics r_k is position, posture and velocity on a body part to another body part. r_k belong to FMFs group $\boldsymbol{r} \in \mathbb{R}^m$, where m is the

number of FMFs, and $\{\boldsymbol{r}\}_l$ denote the l-th component of $\boldsymbol{r}$, that is, $\phi_{\mathrm{fmf},k}(\boldsymbol{x}) = \{\boldsymbol{r}\}_l$. b_k is optimized to minimize error rate ϵ_k at k-th round of learning process. β_k represents that the assigned action occurs or not at a certain frame. If the assigned action occurs, $\beta_k = 1$ and value of the motion feature is no less than the threshold. If the assigned action does not occur, $\beta_k = -1$ and value of the motion feature is no more than the threshold. After all, scope of β_k is $\beta_k = \{+1, -1\}$.

The design of action cues allows action classifiers to select automatically a scalar motion feature r and its threshold β_k in learning process. This learning process can be viewed as the process of feature selection. In addition the parameter given by humans is only the number of action cues contained in each action classifier.

Here lists the merits of our action cue. 1) calculation cost for classification is very small. 2) boosting process automatically selects important motion features for classification. 3) the parameter given by humans is only the number of action cues contained in each action classifier.

Actions can be represented with relationship among multiple body parts and we call the relationship of each action "Action Rule". For example, action rule of *lying* is that *lying* is an action whose height of head and hip is on a line, and action rule of *folding arms* is that *folding arms* is an action both of whose hands cross in front of chest.

Boosting process can obtain action rule of targeted action as action cues in feature selection process. However, FMFs based action cues can only represent action rule as relationship among two body parts, and can't represent action rule as relationship among above two body parts. Therefore, it is often difficult for humans to understand action rule whose action have relationship among body parts as characteristic of the actions. And action classifiers are inefficiently constructed because these contain many action cues. In addition, the classifier using only FMFs based action cues hardly recognize robustly actions whose postures are complicated, such as *folding arms* and *running*.

Accordingly, we introduce combinational motion features (CMFs), which allow action cues to represent relationship among above two body parts, in order to describe briefly action rule for knowledge discovery and construct efficiently action classifiers with a small number of action cues.

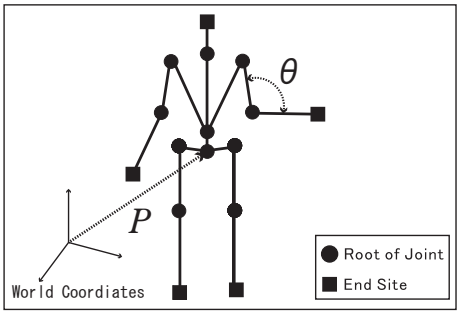

Fig. 4. Measured motion data

C. Action cues with combinational motion features

As stated above, action classifier using only FMFs based action cues is constructed inefficiently with many action cues and often can't recognize action robustly. In addition, action rule obtained in boosting learning process is difficult for humans to understand because the action rule is complicated.

Therefore we introduce combinational motion feature (CMFs) which are the combinations of several FMFs. Combination process utilize various template operations.

For example, we combined FMFs by utilizing $\{\boldsymbol{r}\}_l - \{\boldsymbol{r}\}_{l'}$ $(l \neq l')$ as a template operation and create CMFs. In classification task shown in Fig 5, the classifier using only FMFs based action cues needs many action cues, however the classifier using not only FMFs based action cues but also CMFs based action cues needs only one action cue. Generally, utilization of CMFs based action cues allows classifier to detect critical multiple feature space for classification while retaining small calculation cost for classification. Each CMFs based action cue is described as follows:

$$h_k(\boldsymbol{x}) = \begin{cases} +1, & [\![\beta_k \phi_{\mathrm{cmf},k}(\boldsymbol{x}) \geq b_k]\!] = 1 \\ -1, & \text{otherwise} \end{cases},$$

where $\phi_{\mathrm{cmf},k}$ is the function which extract a scalar CMF from measured motion data $\boldsymbol{x}$. This instance is $\phi_{\mathrm{cmf},k}(\boldsymbol{x}) = \{\boldsymbol{r}\}_l - \{\boldsymbol{r}\}_{l'}$.

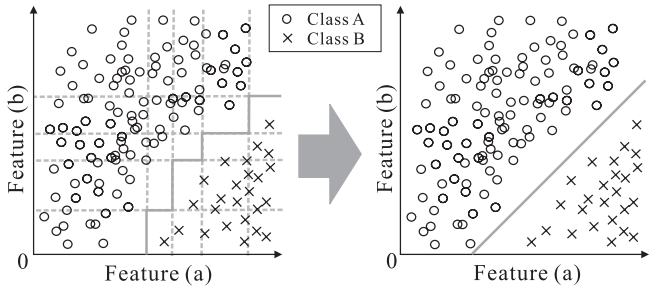

Fig. 5. Example of Utilizing Combinational Motion Features

In order to show usefulness of CMFs based action cues, we quote an example of classification of *folding arms* whose data structure is similar to the data structure in Fig 5. In recognizing *folding arms*, we expect that constructing action cues in view of relationship among both hands and chest is more efficient than constructing independently action cues relating to each body parts. Thus, in view of relationship among both hands and chest, we utilize the action cue that classifies by threshold processing with difference in coordinate values of lateral both hands positions for chest (Fig 6). That is, we utilize the action cue whose template operation CMF $\phi_{\mathrm{cmf},k}$ is $\{\boldsymbol{r}\}_l - \{\boldsymbol{r}\}_{l'}$, where $\{\boldsymbol{r}\}_l$ denotes lateral position of left hand for chest and $\{\boldsymbol{r}\}'_l$ denotes lateral position of right hand for chest. Action rule described with this CMF based in *folding arms* is that *folding arms* is an action both of whose hands cross in front of chest. This action rule is easy for humans to understand. Through above discussion, we expect that CMFs reduce the number of action cues contained in the classifier and describe action rule understood by humans in *folding arms* recognition.

There is a lot of template operations CMF $\phi_{\mathrm{cmf},k}$, such as $\phi_{\mathrm{cmf},k}(\boldsymbol{x}) = \{\boldsymbol{r}\}_l + \{\boldsymbol{r}\}_{l'}$ and $\phi_{\mathrm{cmf},k}(\boldsymbol{x}) = \{\boldsymbol{r}\}_l{}^2 + \{\boldsymbol{r}\}_{l'}{}^2$. However, we should avoid overelaborating the design of $\phi_{\mathrm{cmf},k}$ in consideration of calculation cost for learning process. In addition, we limit CMFs used in this paper to be readable as action rule for knowledge discovery. For example, it is not used as CMF that the combination of position of hand and posture of hip, or the combination of position of hip and velocity of hip.

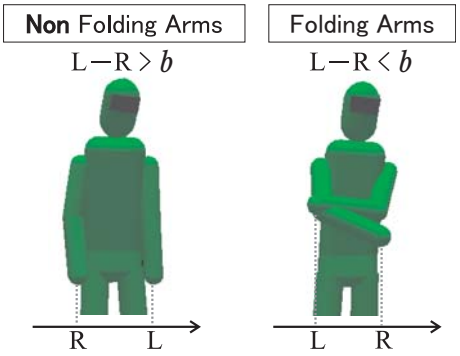

Fig. 6. Example of Combinational Motion Features with Lateral Positions of Both Hands For Chest: L/R is a Lateral Position of Left/Right Hand for Chest

D. Design for new action cues

Our previous action cues difficultly form the appropriate separating boundary in the classification task showed in Fig 7.

In this case, it is necessary to form rectangle separating boundary. Consequently we design a new action cue h_k described as follows:

$$h_k(\boldsymbol{x}) = \begin{cases} +1, & \prod_{m=1}^{M} [\![\gamma_{k,m}\psi_{k,m}(\boldsymbol{x}) \geq c_{k,m}]\!] = 1 \\ -1, & \text{otherwise} \end{cases},$$

where $\gamma_{k,m} = \{+1, -1\}$, and $c_{k,m}$ is threshold of k-th action cue, both parameters are optimized in learning process. As in the case of $\phi_{\mathrm{fmf},k}$ and $\phi_{\mathrm{cmf},k}$, $\psi_{k,m}$ extracts a scalar motion feature from measured motion data $\boldsymbol{x}$. In Fig 7, $M = 2$ and $\psi_{k,1}(\boldsymbol{x}) = \{\boldsymbol{r}\}_l$, $\psi_{k,2}(\boldsymbol{x}) = \{\boldsymbol{r}\}_{l'}$.

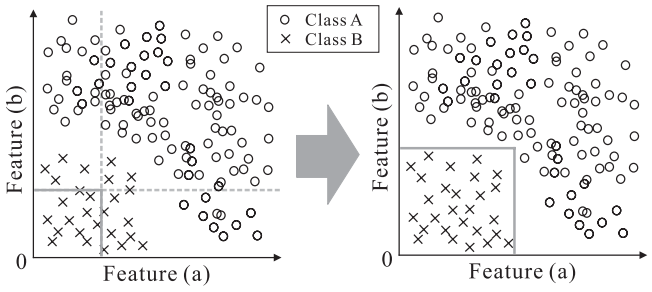

Fig. 7. Example of Utilizing New Action Cue : Rectangle Type

For example, in *running* classification whose structure is similar to the data structure in Fig 8, action classifier is efficiently constructed with the new action cues and recognize *running* robustly. In Fig 8, $M = 2$, $\{\boldsymbol{r}\}_l$ denotes degree of bend of left knee and $\{\boldsymbol{r}\}_{l'}$ denotes degree of bend of right knee. Action rule of *running* is that *running* is an action where both knees greatly bend, then this action rule is easy for humans to understand.

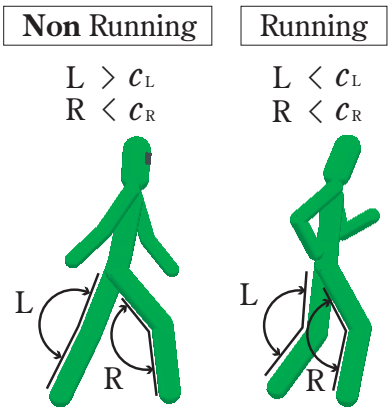

Fig. 8. Example of New Action Cues with Degree of Bend of Both Knees: L/R is a Degree of Bend of Left/Right Knees

As noted in this section, the introduction of CMFs allows us to construct robust classifiers with a few action cues and the action classifier to classify the actions which are not classified robustly by using only FMFs based action cues. In addition, action rule described with CMFs based action cues is briefly.

As noted in this section, the introduction of CMFs allows us to construct robust classifiers with a few action cues and to solve the problem of recognizing the actions which are not classified robustly using only action cues with FMFs. In addition, action rule described with CMFs based action cues is briefly.

IV. Experimental results

This section describes evaluation experiment for the effectiveness of the proposed method. First, we evaluate the proposed method in terms of performance and calculation cost for classification. Second, we demonstrate the effectiveness of the introduction of CMFs by comparing FMFs based action classifier. Finally, we show an example of classification result with the proposed method.

A. Target actions to be classified

We select *walking* and *running* as actions targeted to be classified in the experiments. Daily life actions contain actions without movements such as *lying* and *sitting*, and actions with movements such as *walking* and *running*. However we target only actions with movements in the experiments. This is because actions without movements have innate poses and actions with movements vary according to time and then have not innate poses, and then we expect that actions without movement is easy to be classified compared to classification of actions with movements. Hence, if classifiers based on the proposed method classify actions with movements robustly, actions without movements are also classified robustly.

TABLE II
MOTION DATA FOR EXPERIMENTS

# of frames	Set1	Set2	Set3
total	1079	1094	1043
Walking	318	320	292
Running	233	240	236

B. Candidates for action cues

In the experiments, since targeted actions are gait motions, we mark only lower parts of human's body. Specifically, the number of body parts marked are 7 in total; hip, both feet, both legs, and both thighs.

In view of calculation cost for learning process and for knowledge discovery, the template operations of CMFs $\phi_{\mathrm{cmf},k}(\boldsymbol{x})$ are $(\{\boldsymbol{r}\}_l + \{\boldsymbol{r}\}_{l'})$, $(\{\boldsymbol{r}\}_l - \{\boldsymbol{r}\}_{l'})$, $(\{\boldsymbol{r}\}_l + \{\boldsymbol{r}\}_{l'} + \{\boldsymbol{r}\}_{l''})$, $(\{\boldsymbol{r}\}_l + \{\boldsymbol{r}\}_{l'} - \{\boldsymbol{r}\}_{l''})$, $(\{\boldsymbol{r}\}_l - \{\boldsymbol{r}\}_{l'} + \{\boldsymbol{r}\}_{l''})$, $(\{\boldsymbol{r}\}_l - \{\boldsymbol{r}\}_{l'} - \{\boldsymbol{r}\}_{l''})$, $(\{\boldsymbol{r}\}_l{}^2 + \{\boldsymbol{r}\}_{l'}{}^2)$, $(\{\boldsymbol{r}\}_l{}^2 - \{\boldsymbol{r}\}_{l'}{}^2)$, total 8 patterns. As for $\psi_{k,m}(\boldsymbol{x})$, we utilize only $\{\boldsymbol{r}\}_l$ as $\psi_{k,m}(\boldsymbol{x})$, and set $M = 2$.

As noted above, the number of FMFs group $\boldsymbol{r}$ calculated from measured motion data is 229 ($\boldsymbol{r} \in \mathbb{R}^{229}$), and the number of CMFs constructed from the FMF is 9453. Thus, the number of motion feature candidates ($\phi_{\mathrm{fmf},k}(\boldsymbol{x})$, $\phi_{\mathrm{cmf},k}(\boldsymbol{x})$ and $\psi_{k,m}(\boldsymbol{x})$) is 9682 in learning process of action cues.

C. Motion data set

The measured motion data in the experiments is sequential human motion data fetched by a magnetic motion capturing system. The format of the data file is BVH, a de-facto standard motion file format by Biovision Corporation. A BVH file contains the structure of a human as a linked joint model figure. The model in the BVH of the experiments has total of 36 DOFs and the motion of the figure per frame. The body motion is measured at 30 Hertz.

The actions included in the motion capture files are *walking*, *running*, *stand still* and transition from an action to another action. We annotate motion data per frame with *walking* or with non *walking*, and with *running* or with non *running*.

The motion data is divided into 3 sets and Table II shows the number of frames in each set. Of 3 sets, 2 of 3 sets are utilized as training data and 1 of 3 sets is utilized as test data, and we evaluate the proposed method by cross validation.

D. Evaluation Method

For performance measure, we use F-measure. F-measure is a harmonic average of recall rate and precision rate. R denotes recall rate, P denotes precision rate, and then F-measure can be defined as follows:

$$F = \frac{2RP}{R+P}$$

F-measure is an indicator of classifier's ability to detect all frames of targeted action without mistakes. The bigger F-measure is, the higher recognition performance is.

TABLE III
COMPARATIVE RESULT OF PROPOSED METHOD AND THE SVM-BASED METHOD

F - measure	Walking	Running
Proposed method	93.3%	91.9%
SVM-based method	93.2%	92.5%

E. Experiment for effectiveness of classifier based on boosting

We evaluate the proposed method based on boosting in terms of performance and calculation cost for classification. The classifier based on SVM with Gaussian kernel that classifies by nonlinear separating boundary on feature space is used as comparative method.

The classifier based on SVM uses all FMFs as motion features. The classifier based on the proposed method is constructed by 100 action cues. Table III shows experimental result of the proposed method and the SVM-based method.

Table III tells that the proposed method was as well as the SVM-based method in performance. Moreover, the average of calculation time per frame for classification was about 0.03 milliseconds in the proposed method and about 3 to 4 milliseconds in the SVM-based method. Namely calculation time of the proposed method is much faster than of the SVM-based method. In addition, calculation time of the proposed method is independent of the number of training samples, contrary to this, calculation time of the SVM-based method generally tends to increase in proportion to the number of training samples [14]. Moreover, daily life actions contain many actions. Hence it is important for classifying many actions simultaneously that calculation time of classification is as small as possible.

This experiment yields that the proposed method has classification performance as well as the method based SVM with Gaussian kernel and can classify fast in practical.

F. Experiment for effectiveness of CMFs

In this experiment, we compare the classifier with CMFs to the classifier without CMFs so as to reveal the effectiveness of CMFs. Fig 9 shows transition of classification performance for test data versus the number of action cues when data set 1 and 2 were used as training data.

CMFs enable action classifiers to have high classification performance. And the action classifier with CMFs is most always more robust than the action classifiers without CMFs on the condition that each classifier is constructed with the same number of action cues.

G. Example of classification with the proposed method

Fig 10 shows an example of recognition result with the proposed method and thumbnail of this motion data file. Besides, Fig 10 shows recognition result with the method based on SVM with linear kernel. Calculation cost of the method based on SVM with linear kernel is as small as of the proposed method. Upper thumbnail shows human figures fetched every 0.17 seconds and per 5 frames in targeted sequential motion

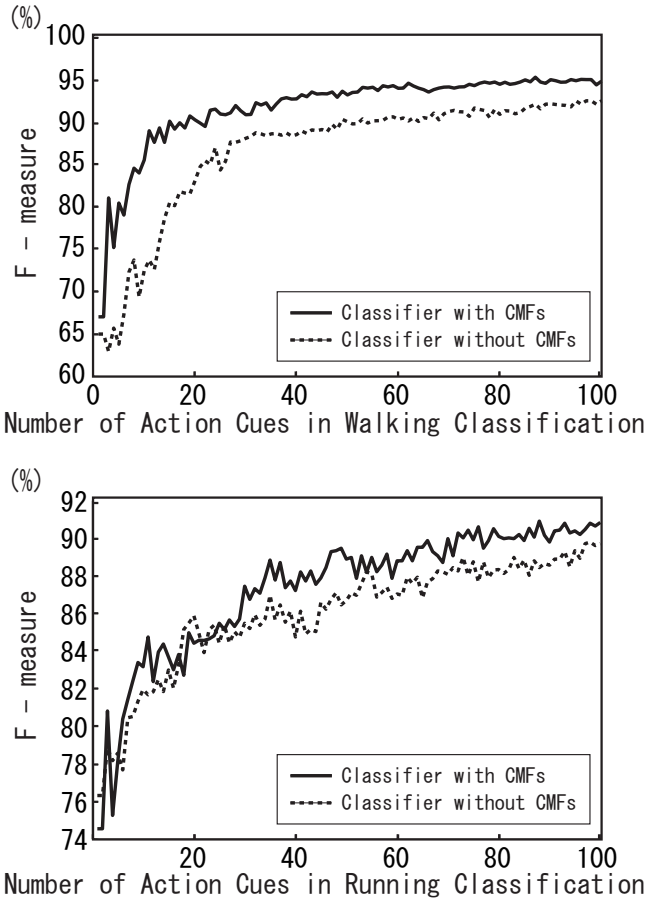

Fig. 9. F-measure for Test Data vs. Number of Action Cues in Walking and Running Classification

data file. Lower left graph shows the recognition result with the proposed method, and lower right graph shows the recognition result with the method based on SVM with linear kernel. In the two graphs, horizontal thick solid lines indicate the estimated action label and vertical lines indicate start or finish of actions.

Fig 10 shows that the proposed method, in spite of not considering interdependencies of output action labels, can recognize actions robustly and catch the shift from an action to another action. Contrary to this, the comparative method can scarcely recognize actions robustly.

V. Conclusion

In this paper, we proposed a method based on boosting for realization of robust online daily life action recognition with small calculation cost. We premised the online recognition system that contains classifiers working independently and in parallel, and these classifiers were constructed based on boosting; an ensemble learning algorithm.

Action cues, which are important for booting algorithm, were designed to classify by threshold processing of a scalar motion feature. This approach allows classifiers to lessen calculation cost for recognition and to select automatically critical motion features. Moreover, the parameter that humans must give an action classifier is only the number of action cues contained in an action classifier with the proposed method. We expect that there are various many methods for automatic optimization of the parameter. However, we cannot decide a conclusive method because it is not necessarily that recogniton performance for test data is well when recognition performance for training data is well. And because required recognition performance depend on an application. Thus, we did not discuss automatic optimization of the parameter in this paper.

We also proposed boosted combinational motion features (CMFs) and exploited boosted CMFs to construct an actions classifier. There are three reasons to propose. First is the action classifiers can recognize robustly actions that can be hardly recognized robustly by utilizing only boosted fundamental motion features (FMFs). Second is that we expect the action classifiers to be constructed efficiently with a small number of action cues. Third is human can understand action rule obtained in the action classifier learning process.

We evaluated the proposed method by applying the method to recognition for gait motion; *walking* and *running*. The motion data was fetched by motion capturing system. In consequence, the action classifiers exploiting boosted CMFs achieved high recognition performance and whose calculation time for recognition was much smaller than of the method based on SVM with Gaussian kernel. In addition, the number of action cues of the classifier with boosted CMFs was less than of the classifier without boosted CMFs while the classifier with boosted CMFs can classify robustly. Besides, classifiers with the proposed method can recognize actions robustly and catch a transform from an action to another action despite the classifiers ignore sequential classification result.

Future work is to propose the method that takes account of interdependencies of output action labels in order to construct recognizer which is strong for noise and lack of motion data.

References

[1] J. Davis and A. Bobick. The representation and recognition of human movement using temporal templates. In *Proc. of CVPR*, pages 928–934, 1997.

[2] J. Yamato et al. Recognizing human action in time-sequential iages using hidden Markov model. In *Proc. of CVPR*, pages 379–385, 1992.

[3] T. Mori et al. Human-like action recognition system using features extracted by human. In *Proc. of IROS*, pages 1214–1220, 2002.

[4] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proc. of the IEEE*, volume 77, pages 257–285, 1989.

[5] B. Schölkopf and A. Smola. *Learning with Kernels*. MIT Press, 2002.

[6] T. Inamura et al. Imitation and primitives symbol acquisition of humanoids by integrated mimesis loop. In *Proc. of ICRA*, pages 4208–4213, 2001.

[7] D. Cao et al. Online motion classification using support vector machine. In *Proc. of ICRA*, volume 3, pages 2291–2296, 2004.

[8] T. Mori et al. Recognition of actions in daily life and its performance adjustment based on support vector learning. *Intl. Journal of Humanoid Robotics*, 1(4):565–583, 2004.

[9] E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Machine Learning*, 36(1-2):105–139, 1999.

[10] Y. Freund. Boosting a weak learning algorithm by majority. In *Proc. of COLT*, pages 202–216, 1990.

[11] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

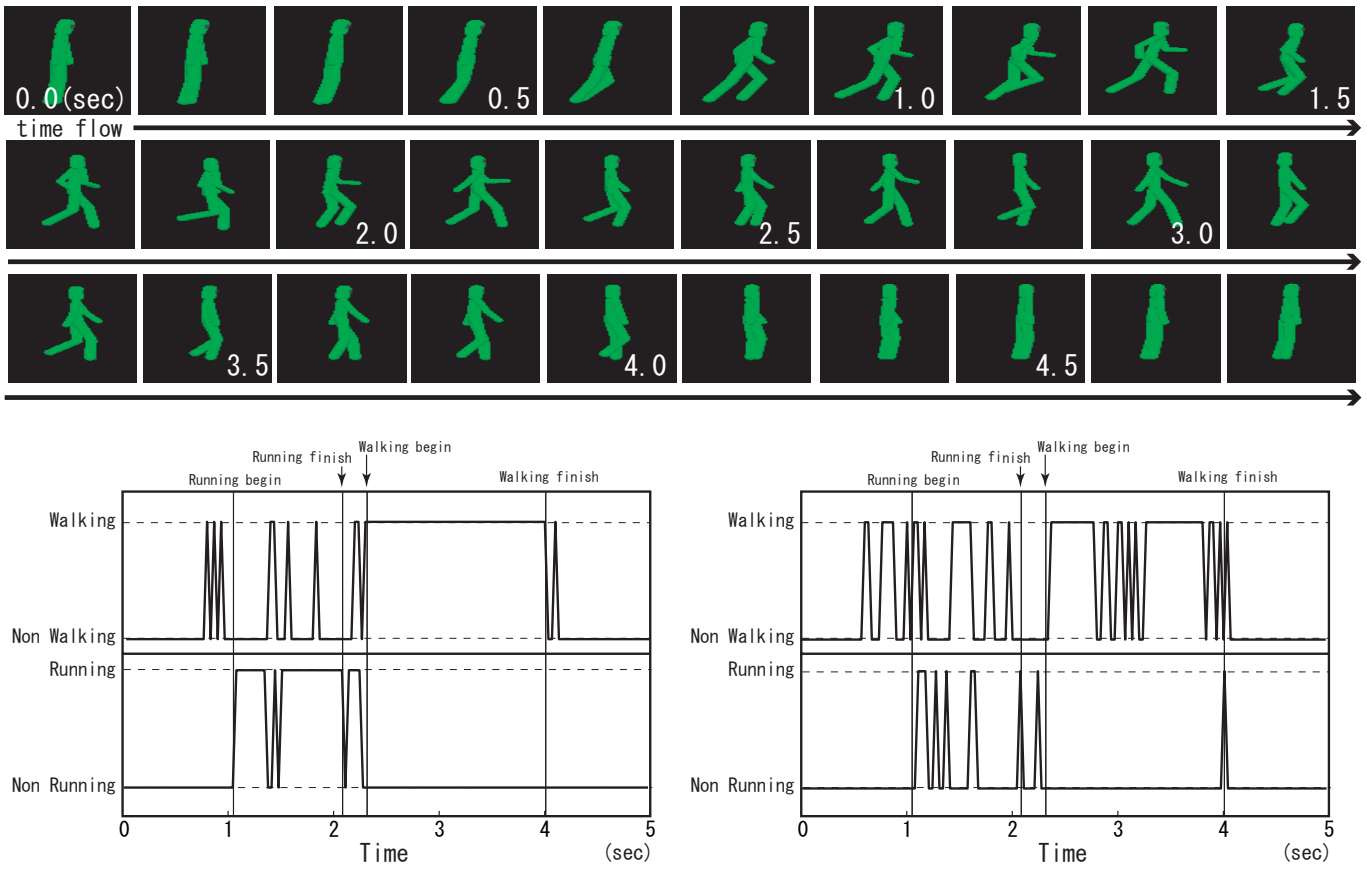

Fig. 10. Thumbnails of Motion Data Example for Testing Proposed Method (Left) and SVM with Linear Kernel based Method (Right)

[12] P. Viola et al. Detecting pedestrians using patterns of motion and appearance. In *Proc. of ICCV*, volume 2, pages 734–741, 2003.

[13] T.Mori et al. Recognition of daily action based on automatic feature selection using adaboost (in japanese). In *Proc. of SICE SI 2006*, pages 545–546, 2006.

[14] O.Chapelle et al. Choosing multiple parameters for suppurt vector machines. *Machine Learning*, 46(1):131–159, 2002.