SVM-based Human Action Recognition and its Remarkable Motion Features Discovery Algorithm

Taketoshi MORI, Masamichi SHIMOSAKA, Tomomasa SATO
Mechano-Informatics, The University of Tokyo tmori@ics.t.u-tokyo.ac.jp

Motivation, Problem Statement, Related Works

This paper proposes a discovery algorithm of knowledge of remarkable motion features in daily life action recognition based on Support Vector Machine. The main characteristics of the proposed method are 1)basic scheme of the algorithm is based on Support Vector Learning and its generalization error, 2)remarkable motion features are discovered in response to kernel parameters optimization through generalization error minimization. Experimental results show that the proposed algorithm makes the recognition system robust and finds remarkable motion features that are intuitive for human..

Recognizing human actions has potential to contribute to intuitive communication between human and machine, such as human-computer interaction, search engine for multi-media databases, or intelligent video editing. It may be applied to design some level of humanoid actions efficiently. It is proper to divide the process of action recognition into the following two phases. The former is to get time series of 3D body motion structurally from some instruments. The latter is to symbolize these kinds of motion to action names. There are many researches which use video image as input. For example, Starner et al. constructed a sign language recognition based on HMM[1], and Wilson et.al made a gesture recognition system[2]. However, usually the main work of such researches is on how to acquire motion robustly, since the time series image processing is still a difficult problem.

Technical Approach

On the other hand, our approach concentrates on the latter part. We use structured motion capture data as input, and focuses on 'recognition' part. As for the recognition method, we developed a daily life action recognizer which recognized daily life action such as walking or sitting based on Support Vector Machine. The main characteristics of the method is to utilize expressions of action described based on human knowledge. This enables aids for the designer of the recognition method to detect remarkable motion features in motion candidates. But, it is often difficult to generate expressions for actions. And, there is some difficulty of selecting or evaluating relevant motion feature from the described expressions for the actions. Therefore, this paper proposes an algorithm which discovers remarkable motion features for a certain action automatically based on many corresponding motion captured data. The algorithm finds out which motion features are important for which action to be recognized.

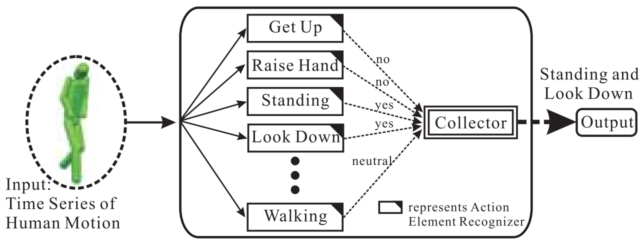

Fig1. Configuration of Recognition Method

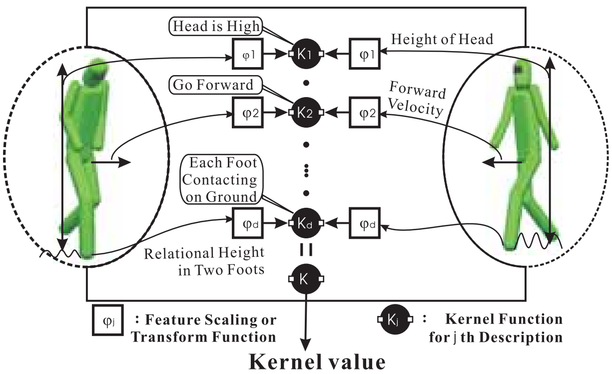

Fig2. Example of 'Walking' Recognizer

The main features of the recognition method are: 1)simultaneous recognition which means the method is multiple parallel action recognition, 2)expressing vagueness of human's action recognition which means the method outputs of not decisive recognition result, 3)utilization of action expressions by human which means the method only uses relevant features of the target actions derived from the expressions[3]. We revised our former fuzzy logic based recognition method[4] to use Support Vector Machine[5] for incorporating learning ability.

The criterion of recognition accuracy in this kind of research is difficult since daily actions have no distinct definition. Thus, the correctness of recognition results are basically based on human's judgment in this research. We defined the correctness of a recognition result is to be generated by humans from moment to moment with some motion as decision result whether one specific action occurs or not. This means, multiple action can be recognized at the same time (For example: Running while waving hands).

Figure 1 shows the processing configuration of the method. To realize simultaneous recognition, the method contains multiple recognition processes, each of which is assigned to a certain action's recognition. This primitive recognition process for one action is called as Action Element Recognizer in this research. The process of each AER runs in parallel with other AERs. The method collects the results of all processes and outputs the results of each recognition process per frame. New action can be recognized simply by adding the corresponding recognizer to the method. This means, the result of one AER process is independent of the others. An AER outputs the recognition result whether the assigned action occurs or not per frame in sync with the input motion. The output consists of multiple classes which represent not only decisive but also inexplicit result. Concretely speaking, the number of the class category is 3, yes, no and neutral. In this research, this each AER's classifier is implemented using Kernel classifier. Figure 2 shows an example recognizer which uses several kernel classifiers, which are denoted *Ki* in the figure. The mapping between the input and the output of the classifier in the AER *f* can be written as

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{l} a_i y_i K(x, x_i) + b\right),$$

where *b* depicts the offset and the function sgn() is a step function where the output is 1 if input is more than 0, and otherwize -1. The learning process in the classifier of AER gets the co-efficiencies(***a*** and *b*) from the training data. Support Vector Machine is utilized as learning method for it. Support Vector Machine has such advantages as good model selection, well regularization and small computational resources. The kernel value in the AER classifier is as the product of all the kernel values corresponding to the similarity in each target motion. Radial Basis Function is used as the kernel for target motion.

Results

The main result of this research is that it treat the sensitivity of the kernel value with respect to change of input as relevance of selection of motion features for recognition. It means, the smaller sensitivity of the kernel value might be desired in the case of less importance for recognition. As for RBF kernel, the remarkable input feature requires smaller variance in relevant input feature than irrelevant. If it is required to know which input feature is relevant, human judges which input motion is remarkable by observing the kernel parameters. In this research, the kernel parameters optimization which adjusts kernel sensitivity is used in order to discover remarkable motion features and to optimize recognition performance.

The generalization error is utilized as the indicator of the kernel parameter optimization. To optimize kernel parameter effectively, the gradient descent algorithm is used. The gradient of the co-efficiency ***a*** is calculated based on the fact that the relation between input and output of SVM can be written only by support vectors.

Fig3. Target 18 Actions

1	Relative horizontal position of right hand to left	2	Sum of distance between hands and body
3	Mean speed of hands	4	Bentness of Hips
5	Mean speed of hips(1)	6	Height of hips
7	Upper direction of head from hips	8	Distance between hips and foots
9	Upper direction of hips from foots	10	Horizontal orientation of head from hips
11	Upper orientation of head from hips	12	Upper orientation of head from ground
13	Mean height of head(1)	14	Height of hips
15	Upper orientation of hips	16	Horizontal orientation of hips
17	Upper direction of head from left hand	18	Upper direction of head from right hand
19	Upper direction of hips from left knee	20	Upper direction of hips from right knee
21	Height of left hand	22	Height of right hand
23	Relative height of left hand from hips	24	Relative height of right hand from hips
25	Mean upper velocity of left hand	26	Mean upper velocity of right hand
27	Highest relative height of hands from head	28	Mean height of head(2)
29	Mean speed of hips(2)	30	Mean speed of left foot
31	Mean speed of right foot	32	Speed of Rotation of hips in vertical axis

Table1. Motion Information List

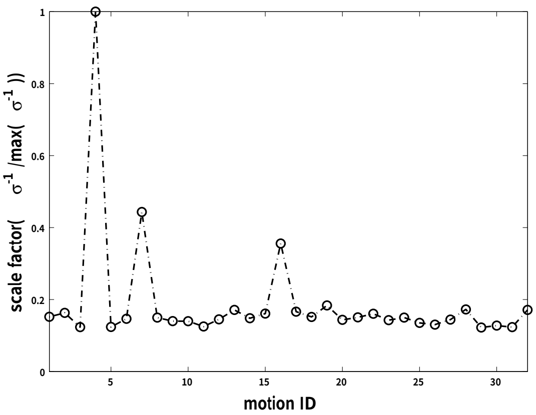

Fig4. Result of Standing

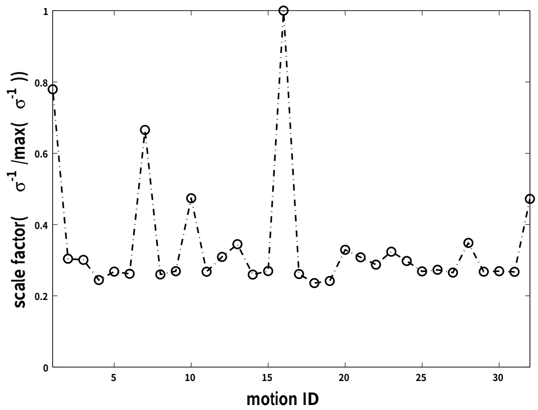

Fig5 Result of Lying of Side

Experiments

Standing, Folding arms etc. from ICS Action Database[6] are used to evaluated the performance of the proposed method. The database is a collection of motion data with full reference action name labels. The format of motion is BVH which is de-facto standard computer graphics motion format. A BVH fil contains the structure of a human as a linked joint model and the time-series motion data of the joint model per frame(usually 30Hz). Total of 18 actions are used(Fig3). The motion information used as feature candidates are listed in Table 1. For the experiments, 125 BVH files and 2250 reference answer action label files are used.

Experimental Snapshots

As the recognition accuracy result, each target action achieves 80[%]. Especially, the accurate rate in 14/18 action names are larger than than 90[%].In almost all the target actions, the accuracy gained by the optimized kernel parameters is better than the case of the initial kernel parameters.

Figure 4 and 5 shows the relative ratio of the kernel parameters after the procedure of the proposed algorithm in the case of "Standing" and "Lying on side", respectively. In each figure, the number on the horizontal axis corresponds to the number in Table 1. The vertical axis shows the normalized inverse of the variances whose maximum value is 1. In the scheme of the proposed detection method, the larger value in the vertical indicates more relevant motion feature, because smaller kernel parameter σ makes kernel more sensitive. In the the case of "Standing", inverse of the variance corresponds to bent of hips is largest. Next, the orientation of the upper body is detected, and horizontal posture of hips is detected as the third relevant motion feature. This result fit with human intuition.

Conclusion

This paper proposes an algorithm for discovering remarkable motion features from candidates of motion features in human daily life action recognition based on Support Vector Machine. This discovery algorithm is based on kernel parameters optimization as minimization of generalization error. The experimental result for performance evaluation shows that the accuracy of recognition achieves high enough, in addition, the performance after the optimization of the kernel parameters is better than the case of the initial settings. It is also shown that the relative importance values which corresponds to inverse of the variances fit with human intuition.

References

[1] T. Starner and A. Pentland. Visual Recognition of American Sign Language Using Hidden Markov Models. In International Workshop on Automatic Face and Gesture Recognition. pp.189-194, 1995.

[2] A. Wilson and A. Bobick. Parametric Hidden Markov Models for Gesture Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.21, No.9, pp.884-900, 1999.

[3] M. Shimosaka et al. "Recognition of Human Daily Life Action and Its Performance Adjustment based on Support Vector Learning". In CD-ROM of Third IEEE Intl. Conf. on Humanoid Robots, 2003.

[4] T. Mori, K. Tsujioka. Human-Like Action Recognition System on Whole Body Motion-Captured File. In Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.2066-2073, 2001.

[5] J. Platt. "Probabilities for SV Machines". In Advances in Large Margin Classifiers, pp. 61–74. MIT Press, 1999.

[6] ICS Action Database. http://www.ics.t.u-tokyo.ac.jp/action/, 2003-2004.