Support for Motor Learning by Visualizing the Similarity of Sports Form
Examining Effective Image Features in Back Hip Circle Videos of Children

Ayumi Matsumoto[1], Dan Mikami[1], Harumi Kawamura[1], Akifumi Kijima[2] and Akira Kojima[1]
[1]*NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation, Kanagawa, Japan*
[2]*Graduate School of Education, University of Yamanashi, Yamanashi, Japan*

1 OBJECTIVES

Video feedback is an effective tool in the field of motor learning. Video motion analysis software such as the Dartfish software (Dartfish, 1997) has been introduced into athletes training programs as well as school physical education and rehabilitation programs. Dartfish aims to support coaching and learning by enabling coaches and athletes to view the motion flow in sports videos and to superimpose multiple forms on it. This software is very useful as a means of incorporating image viewing into coaching and practice methods, but it requires specialized knowledge of the sport to which it is being applied.

The goal of our study is to propose systems that can support motor learning and coaching for users (or teachers) without requiring any detailed knowledge of sports. We propose a system that can visualize similarities in "form" as the term applies to sports (e.g., batting form, hurdling form) and an instruction method that will automatically be suitable for the person using it. As a step toward this goal, in this paper we propose a framework for the similarity visualization that is based on similarities in form and optimal image features for classifying similarities in target forms in sports actions that require different lengths of time to perform.

We expect that visualizing similarities in form in this manner will be effective in two ways. First, through a form classification process it will provide instruction methods that are suitable for the groups that use it. Second, it will allow individual users to assess and evaluate their own form by comparing it with others in the same group.

2 METHODS

2.1 Framework

Figure 1 shows an example of the visualizing of similarities in form we assume. The video is mapped in 2D (or 3D) on the basis of image features, and put into a certain class involving the use of a supervised or unsupervised method. This type of similarity visualization may be able to help teachers determine teaching methods for classes in advance and provide guidance that will suit individual users. Figure 2 shows the work flow of the proposed similarity visualization process. First, it gets the image features from the video. Then, it uses the features to calculate similarities. Finally, it classifies the similarities into any number of classes and displays the classification results and the instruction method that is suitable for the group.

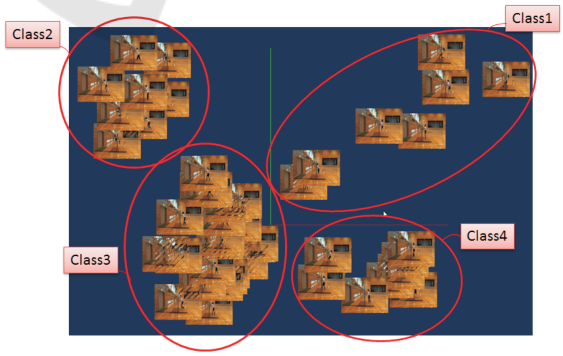


Figure 1: Example of visualizing similarities in sports form.

2.2 Image Features

Appropriate image features have not been studied yet when performing sports form classification. Accordingly, we examined image features that could be effectively applied to form similarity visualization. In this paper, we classify forms using image features that are often used in motion recognition, and describe the improvements we have achieved.

Figure 3(a) shows an image obtained with MHI (Motion History Image), which is one of the image representation methods in which past images are incorporated into a single image (Davis and Bobick, 1997). Figure 3(b) shows one obtained with Bag of

Matsumoto A., Mikami D., Kawamura H., Kijima A. and Kojima A..
Support for Motor Learning by Visualizing the Similarity of Sports Form - Examining Effective Image Features in Back Hip Circle Videos of Children.

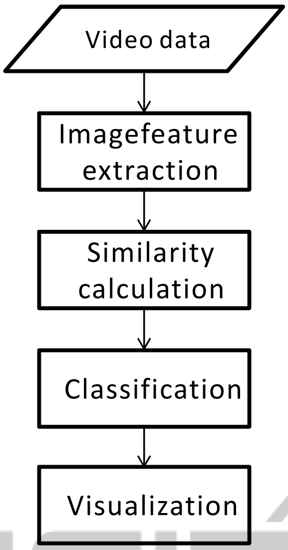


Figure 2: Work flow of proposed similarity visualization process. First, image features are obtained from the video. Then, the features are used to calculate similarities. Finally, the similarities are classified into any number of classes and the classification results and the instruction method that is suitable for the group are displayed.

Video Words (BoVW), which is an image descriptor vector-quantized for time-spatial local features called "cuboids" (Dollar et al., 2005).

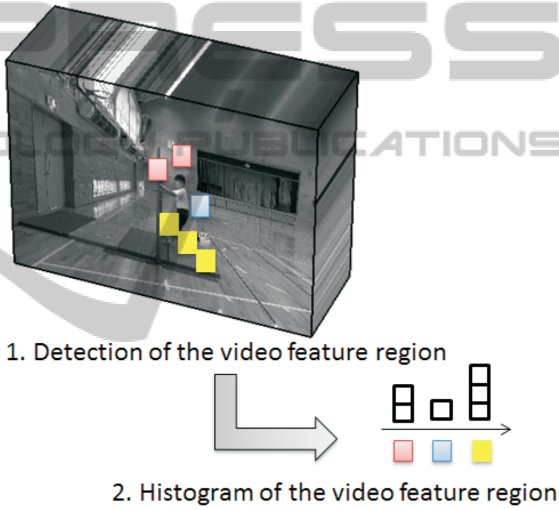
We extended the BoVW concept to include the sequence order information for form, i.e., whether the feature is in the first or second half of the motion sequence. We obtain the new feature by processing using two types of time windows (split window and sliding window) for BoVW . For a video of N frames, a split window divides videos into N/4 frames without overlapping, calculates every BoVW window, and then joins them. A sliding window divides videos into N/4 frames with N/8 overlapping frames.

2.3 Experimental Setting

In this experiment, we classified forms for back hip circle videos of primary schoolchildren. To determine how many effective image features would be needed for form classification, we compared the classification results with four kinds of image features.

The classification targets were back hip circle videos of 178 children in the six grades of elementary school. Video resolution was 1440 x 1080 pixels, the frame capture rate was 30 fps, and each video had length from 150 to 250 frames. An expert assessed the video performances by assigning one of four classes to them: form0false (suspension form in which arm power is used to rotate the body, failure), form1false (warp form in which the person leans back to rotate

(a) Motion History Image (MHI).



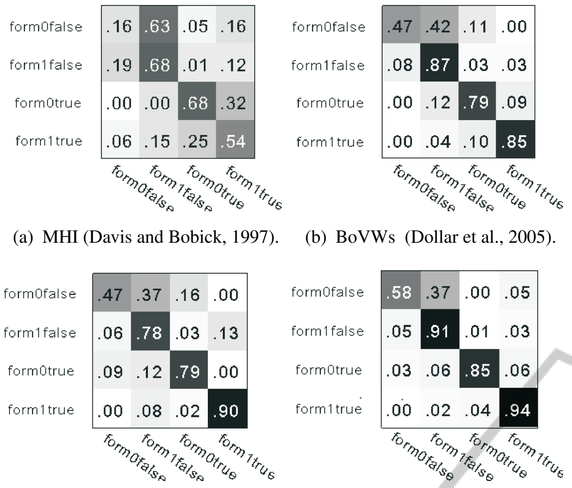
(b) Bag of Video Words (BoVWs).

Figure 3: Video feature.

the body by retroaction, failure), form0true (suspension form, success), and form1true (warp form, success). LDA (Linear Discriminant Analysis) which is the supervised learning technique were used to classified. Classification accuracy of the four image features was determined by dividing the data obtained into test data and learning data.

3 RESULTS

Figure 4 shows a confusion matrix of classification accuracy for back hip circle videos. A confusion matrix is a table summarizing classification results. Rows show the actual classes and columns show the classes that had been predicted. The classification accuracy is higher the closer the number is to 1. In the



(a) MHI (Davis and Bobick, 1997). (b) BoVWs (Dollar et al., 2005). (c) BoVWs (Split window). (d) BoVWs (Slide window).

Figure 4: Confusion matrix of classification accuracy. This is a table summarizing the classification results. Rows show actual classes and columns show classes that had been predicted. The classification accuracy is higher the closer the number is to 1.

figure, 4(a) shows the classification accuracy of MHI, which was the lowest of the methods compared. It is assumed that the reason for the low accuracy is that the history had been overwritten. For the BoVW shown in Fig. 4(b)-4(d), the highest recognition rates were obtained with the sliding window (Fig. 4(d)).

For general classification problems (walking, running, waving, etc.), good classification results may be obtained even if BoVW ignores the order information. However, the results shown in the figure suggest the order information is important in a more detailed differentiation form classification. The classification accuracy we obtained with the split window (Fig. 4(c)) was not very much better than that of conventional BoVW (Fig. 4(b)). We consider that the reason for this is the variation in length of the personal back hip circle videos.

4 CONCLUSION

We proposed a framework for similarity visualization that is based on similarities in form and optimal image features for classifying similarities in target forms in sports actions that require different lengths of time to perform. We performed form classification experiments with the aim of determining appropriate image features in order to visualize form similarities. The highest classification precision was obtained with BoVW by applying a sliding window to back hip circle videos. In future work, we will aim to develop a method that will improve classification accuracy by enabling classification results that reflect users' intentions to be re-learned. Another task would be to evaluate our method's effectiveness in sports training in actual sports applications.

REFERENCES

Dartfish (1997). Dartfish software. http://www.dartfish.com/en/index.htm.

Davis, J. W. and Bobick, A. F. (1997). The representation and recognition of human movement using temporal templates. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, CVPR '97, pages 928–934, Washington, DC, USA. IEEE Computer Society.

Dollar, P., Rabaud, V., Cottrell, G., and Belongie, S. (2005). Behavior recognition via sparse spatio-temporal features. In *Proceedings of the 14th International Conference on Computer Communications and Networks*, ICCCN '05, pages 65–72, Washington, DC, USA. IEEE Computer Society.