LOGOS OF HUMAN ACTIONS

Jian Wang[1], *Wuzhenni Hu*[2], *Zhiling Wang*[1], *Muhammad Sarfaraz Malik*[1], *Zonghai Chen*[1]

[1]Department of Automation, [2]Department of Computer Science
University of Science and Technology of China, China

ABSTRACT

The video sequence which contains certain human action is considered as a spatio-temporal volume. There exists certain characteristic signature in appropriately selected spatio-temporal slice of the video sequence. By using these discriminative signatures which we call “human action logos”, new approaches are proposed for period detection and action recognition. Algorithm performance is evaluated under eight typical human actions. Preliminary experiments have shown promising results.

Index Terms— Human action recognition, human action logos, spatio-temporal slice, period detection

1. INTRODUCTION

Human action recognition is becoming a very hot research area in computer vision with many important applications, such as smart video surveillance, content-based video indexing and human-computer interfaces. Many approaches have been proposed [1], such as action templates based approaches, spatio-temporal interest points based approaches [2], and key frames based approaches. However, we found that some simple actions like waving-one-hand, waving-two-hands, jumping-jack, jumping-in-place-on-two-legs, kicking, walking, etc. can be recognized quickly in a different “frame”—a spatio-temporal slice. Considering the given video containing certain action as a cuboid, cutting it up at some height along with the *t*-axis, we can find some exquisite texture. Different actions have different textures, i.e. different textures represent different actions as various logos are on behalf of corresponding organizations. We call the special textures of human actions as “human action logos” in this paper. We use the logos of actions for human action recognition. Fig.1 illustrates logos of enterprises and logos of human actions.

In a way, our approach can be considered as a template based approach, or a key-frame based approach, or a spatio-temporal interest points based approach. Action logos are also some kinds of templates. Conventional key-frame based approaches act in the frame plane while the proposed approach acts in a special “frame”. Compared to spatio-temporal interest points based approaches, our approach uses the 2D information of the interest points. The step of interest points detection is done by ourselves.

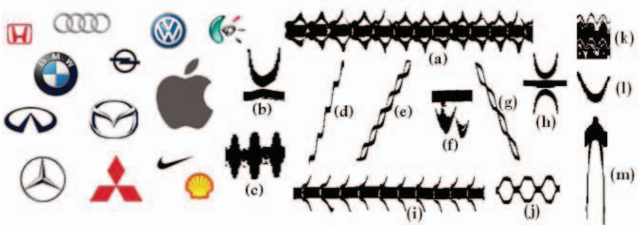


Fig.1: Logos of firms and human actions (a and i: hand clapping; b: waving-one-hand; c and k: jumping-in-place; d: jumping-forward-on-two-legs; e: walking; f: boxing; g: galloping-sideways; h: waving-two-hands; j: jumping-jack; l: bending; m: kicking).

Spatio-temporal slices have attracted attention in some specialized problem domains. They were first used in the context of regularizing structure from motion when the camera motion stayed unchanging [3]. Niyogi and Adelson [4, 5] used the periodic twisted-pattern which is generated by gait in spatio-temporal slices for gait analysis. Liu and Picard [6] analyzed the periodicity of the patterns. Ricquebourg and Bouthemy [7] utilized the twisted-pattern to recognize human pedestrians from vehicles. Ran proposed a computational model for this pattern [8] and approaches for pedestrian segmentation with body part labeling and carrying condition detection [9]. Jacobs and Dixon et al. [10] tracked cars in structured scenes in spatio-temporal sheets. Apostoloff and Fitzgibbon [11] used spatio-temporal T-junctions for segment people from videos. Ngo and Pong et al. [12] characterized and segmented the content of video through spatio-temporal slices. Criminisi and Kang et al. [13] proposed method based on spatio-temporal slices to reason about specular reflection.

The next section lays out action logos, algorithms for human action and period detection. Section 3 describes experiments in which logos based approaches were applied. In Section 4 conclusions and future researches are given.

2. HUMAN ACTION RECOGNITION BASED ON LOGOS OF ACTIONS

2.1 Logos of actions

Fig.2 shows the kind of *yt*-patterns we aim at using for human action recognition. It has been observed that

978-1-4673-0046-9/12/$26.00 ©2012 IEEE

articulated motions reveal such characteristic signatures in appropriately selected *yt*-slices of the video sequence considered as a spatio-temporal volume *xyt*. In Fig.3~Fig.6 and Fig.10 the image on left is a sample image of a video and the right one is a selected slice which contains the logo of the action. We omit the coordinates in Fig.3~Fig.6 and Fig.10 because they share the same coordinates in Fig.2. Data of Fig.2, Fig.3, Fig.9 and Fig.10 are from Weizmann human action dataset [14] and data of Fig.4 and Fig.5 are from KTH human action dataset [15]. Data of Fig.6 are from self-collected dataset.

In the video of Fig.2 a woman is jumping-jack. Logo of jumping-jack at the height of nose is a hyperbolic curve with two little rings while logo of waving-two-hands at the same height is a hyperbolic curve with a little wide line (Fig.6c) for the reason that the height of person jumping-jack is changing but the height of person handwaving is constant.

Another logo of jumping-jack is a hexagon (Fig.6g) which is in the slice at the height of ankle. At the same height, logo of walking is a twisted-pattern (Fig.1e). Logos of galloping-sideways (Fig.3b) and jumping-forward-on-two-legs (Fig.3c) are similar to the twisted-pattern but they have a little bit difference. The intersection part of the twisted-pattern of galloping-sideways is longer than that of walking. The pattern of jumping-forward-on-two-legs is like half of the twisted-pattern for the reason that this action does not own a width-change-mode which exists in walking, galloping-sideways, and running. Logo of kicking at the height of ankle is two vertical lines with a gibbous hat.

At the height of nose, logo of waving-one-hand is a half heyperbolic curve with a little wide line (Fig.6a and Fig.6b) while logo of bending is only a half heyperbolic curve (Fig.6e). Logo of boxing is two arrows (Fig.6d). And logo of jumping is a cross (Fig.6f and Fig.6g). Actually the most distinctive logo of jumping is the “sinusoidal cloud” in the *xt*-slice (Fig.3a) which has been presented in [16]. The logos of walking, jogging and running are all lines but they are different in slope. According to the confusion matrices of many literatures, we can see that most approaches, e.g. approach in [15], are confused by walking, jogging and running because these three actions are similar to each other extremely. In this paper, we think that the most discriminative difference of the three actions is the velocity. Thus if the signature in *yt*-slice at the height of nose is a line and the signature at the height of ankle is a twisted pattern (the twisted pattern ensures that the moving subject is a person [7]), high slope represents the running action, middle slope represents the jogging action and low slope represents the walking action.

Logo of hand-clapping at the height of chest is like a body segment of a centipede (Fig.4).

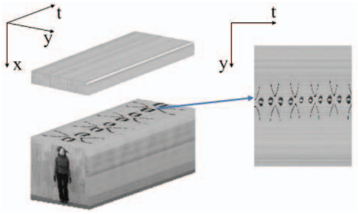
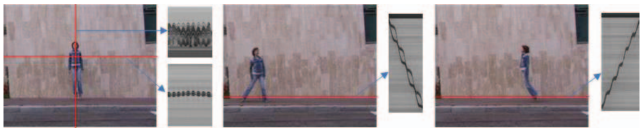


Fig.2: Spatio-temporal slice.



(a) (b) (c)

Fig.3: Appropriately selected spatio-temporal slices of jumping-in-place-on-two-legs (a), galloping-sideways (b), jumping-forward-on-two-legs (c).

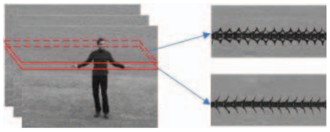


Fig.4: Selected two spatio-temporal slices of hand-clapping.

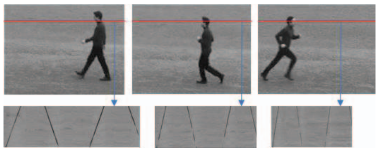
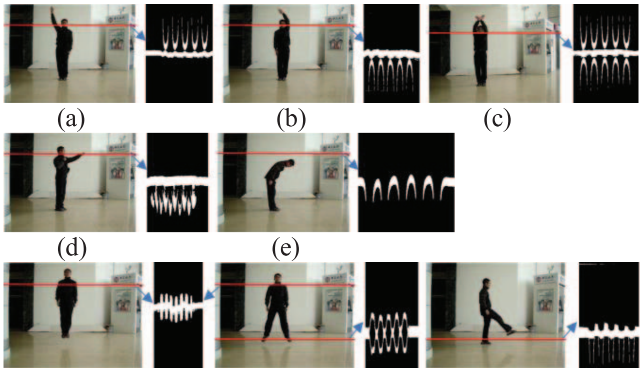


Fig.5: Spatio-temporal slices of walking (left), jogging (middle), and running (right) selected at the height of nose.



(a) (b) (c)

(d) (e)

(f) (g) (h)

Fig.6: Spatio-temporal slices of waving-right-hand (a), waving-left-hand (b), waving-two-hands (c), boxing (d), bending (e), jumping-in-place-on-two-legs (f), jumping-jack-without-hands-waving (g), and kicking (h). Slices of (a) ~ (f) and the left slice in (g) are selected at the same height of nose. The right slice in (g) and slice of (h) are selected at the same height of ankle.

2.2 Human action recognition

The proposed algorithm for human action recognition has following steps: 1) establishing an action logo dataset in which each logo model includes its visual representation, the action(s) and the place(s) where the logo can be found in the action volume(s); 2) pre-processing to the given video such as background modeling, silhouette extraction, height normalization; and finally 3) finding logo(s)—if we find action *A*'s characteristic logo in some spatio-temporal slice of the given video and only action *A* can generate the logo, the action in the given video is classified as *A*; if there are more than one action can generate the logo, the search scale of the action target is narrowed down to these actions and further classification will rely on the distinctive logos of these actions.

2.3 Period detection

We propose a period detection approach based on the logo arrangement in the spatio-temporal slice. We can observe from Fig.7a that the period of the logo is the period of the hand-waving action. In order to detect the period of the logo, we 1) compute the differentials between the uppermost and lowest nonzero pixel' *y* coordinate values (Fig.7b); 2) compute the autocorrelation signal of the differential signal (Fig.7c); 3) compute the first-order derivative to find peak position of the autocorrelation by seeking the positive-to-negative zero-crossing points; and finally 4) estimate the real period as the average distance between each pair of consecutive peaks (e.g. hand-clapping) or major peaks (e.g. hand-waving) [17].

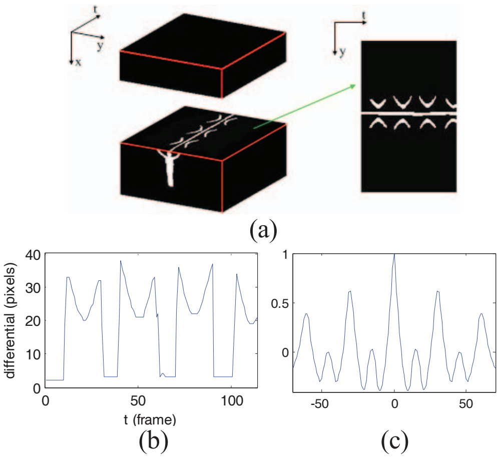


Fig.7: Period detection based on spatio-temporal slice. (a) Generation of the *yt*-slice. (b) Differential signal. (c) Autocorrelation signal of the differential signal.

3. EXPERIMENTS

3.1 Dataset

Experiments were carried on a dataset which contains eight actions—waving-left-hand, waving-right-hand, waving-two-hands, jumping-jack-without-hands-waving, boxing, kicking, bending, jumping-in-place-on-two-legs. A digital camera (Sony W300) fixed on a tripod was used to capture these actions. The subject performed each action for five periods. The action sequences were captured at a rate of 30 fps and the resolution is 480×640. Some sample images of these actions are shown in Fig.6.

3.2 Design of algorithm and the experimental results

We detect period based on the approach proposed in Section 2.3, normalize period, and segment the videos to clips. Each clip contains only one action of a period. After this step we can get 40 video clips which will be divided into two disjoint sets, and one containing eight clips of the different eight actions will be used to establish the logo dataset and the other one containing 32 clips will be taken as testing data. According to the variation of the logos of the actions in the experimental dataset, we design the algorithm as follows: 1) establishing the action logo dataset (Fig.8); here we denote actions of Fig.6a~h as action1~action8 in turn, and model1~model8 indicate action1~action8 successively; 2) after given a video clip, selecting the *yt*-slice at the height of nose, computing the distances of the *yt*-slice to model1~model6 which sizes are $a \times b$ all the same by

$$\min_{m,n} \left\| \text{model}i(1:a,1:b) - yt\_\text{slice}(m:m+a-1,n:n+b-1) \right\|_{\text{Euclid}} \quad (1)$$

in which *a*, *b*>0 and (*m*+*a*-1)×(*n*+*b*-1) is smaller than the size of the *yt*-slice. If the minimum of the six distance values is smaller than a threshold, we apply the Nearest Neighbor classifier for recognition; note that if the distance to model6 is the minimum i.e. action in the clip is classified as class 6, we still do not know whether the action is action6—jumping-in-place-on-two-legs or action7—jumping-jack-without-hands-waving because both of actions can generate model6; and finally 3) selecting the *yt*-slices of video clips which are unclassified or of class 6 classified by step 2) at the height of ankle and computing the distances of each *yt*-slice to model7 and model8 by (1), classifying them with the following rules: the distance to model6 is small and the distances to model7 and model8 are large—action6; the distance to model7/model8 is the smallest—action7/action8 respectively.

The proposed algorithm on the dataset reached the correct rate of 100% and operated in real time.

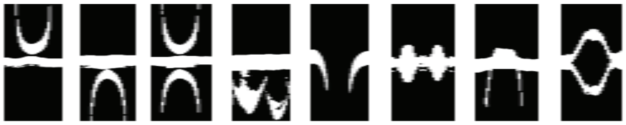


Fig.8: Logo dataset (from left to right are model1~model8).

4. CONCLUSIONS AND FUTURE RESEARCHES

We propose approaches based on logos of human actions for period detection and action recognition. The algorithm performed well on our small dataset containing eight actions. We summarize the main advantages of logos based action recognition: 1) it solves the classification problem by the style hitting the nail on the head; 2) it has a low computational complexity; 3) as long as the action itself is not occluded, the action can be recognized; 4) it can count the occurrence of the action.

What we still need to spare efforts focuses on the following aspects: 1) only one dimensional information is used to detect period in Section 2.3; multidimensional information will perform better; 2) the distance measurement based on Euclidian distance is rough; there exists affine transformation in the signatures of the same action performed by different styles (Fig.9); local invariant feature detectors and descriptors [18] used in image classification and texture recognition may solve the problem; the area “logo recognition” may give hints to us; 3) an advance algorithm can be proposed to remove the pre-processing including background modeling and silhouette extraction and may be directly applied to recognize actions of KTH human action dataset; 4) a future approach can solve slight occlusions which happen at the action itself e.g. vertical occlusion (Fig.10a) or horizontal occlusion (Fig.10b); although the twisted-pattern breaks in the *yt*-slice (Fig.10a, Fig.10b), there exist strong evidences in the slice indicating that the action is walking; further we can utilize the rest twisted-pattern to “remove” the occlusions; the mathematical theory of Geometry Group Theory may be taken to describe the computational model for these action logos as in [8] and this computational model may handle occlusions; 5) similar to human walking, dog running has a double-twisted-pattern in *yt*-slice (Fig.10c); an approach may be proposed to recognize human walking through removing even recognizing the dog’s pattern; because in the spatio-temporal slice the task of counting the dog’s steps becomes easier, maybe it can be used in dog training; and finally 6) the automatic selection of *yt*-slices or *xt*-slices should be researched; maybe entropy based approach will work; approaches based on key frames in human action recognition may give cues.

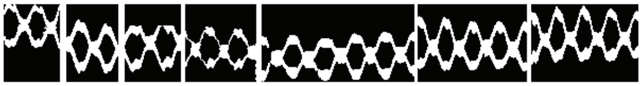
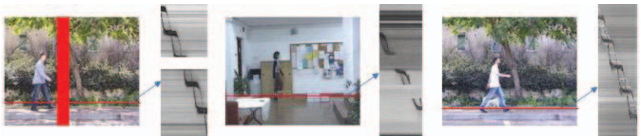


Fig.9: Signatures of jumping-jack performed by different people.



(a) (b) (c)

Fig.10: Some problems to be solved: (a) and (b) occlusion, (c) interference by a dog.

ACKNOWLEDGEMENT

The authors would like to thank the National Natural Science Foundation of China (Grant No. 61005091) and the Specialized Research Fund for the Doctoral Program of Higher Education (Grant No. 20093402110014) for supporting this study.

6. REFERENCES

[1] D. Weinland, R. Ronfard, and E. Boyer, "A survey of vision-based methods for action representation, segmentation and recognition," *Computer Vision and Image Understanding,* vol. 115, no. 2, pp. 224-241, 2010.

[2] H. Wang, M. M. Ullah, A. Klaser, *et al.*, "Evaluation of local spatio-temporal features for action recognition," *British Machine Vision Conference*, pp. 1-11, 2009.

[3] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion," *International Journal of Computer Vision,* vol. 1, no. 1, pp. 7-55, 1987.

[4] S. A. Niyogi and E. H. Adelson, "Analyzing and recognizing walking figures in XYT," *Computer Vision and Pattern Recognition*, pp. 469-474, 1994.

[5] S. A. Niyogi and E. H. Adelson, "Analyzing gait with spatiotemporal surfaces," *IEEE Workshop Nonrigid Articulated Motion*, pp. 64-69, 1994.

[6] F. Liu and R. W. Picard, "Finding periodicity in space and time," *International Conference on Computer Vision*, pp. 376-383, 1998.

[7] Y. Ricquebourg and P. Bouthemy, "Real-time tracking of moving persons by exploiting spatio-temporal image slices," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 797-808, 2000.

[8] Y. Ran, R. Chellappa, and Q. Zheng, "Finding Gait in Space and Time," *International Conference on Pattern Recognition*, pp. 586-589, 2006.

[9] Y. Ran, Q. Zheng, R. Chellappa, *et al.*, "Applications of a simple characterization of human gait in surveillance," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 40, no. 4, pp. 1009-1020, 2010.

[10] N. Jacobs, M. Dixon, S. Satkin, *et al.*, "Efficient tracking of many objects in structured environments," *IEEE 12th International Conference on Computer Vision Workshops*, pp. 1161-1168, 2009.

[11] N. Apostoloff and A. Fitzgibbon, "Automatic video segmentation using spatiotemporal T-junctions," *Proceedings of the British Machine Vision Conference*, 2006.

[12] C. W. Ngo, T. C. Pong, and H. J. Zhang, "Motion analysis and segmentation through spatio-temporal slices processing," *IEEE Transactions on Image Processing,* vol. 12, no. 3, pp. 341-355, 2003.

[13] A. Criminisi, S. B. Kang, R. Swaminathan, *et al.*, "Extracting layers and analyzing their specular properties using epipolar-plane-image analysis," *Computer Vision and Image Understanding,* vol. 97, no. 1, pp. 51-85, 2005.

[14] L. Gorelick, M. Blank, E. Shechtman, *et al.*, "Actions as space-time shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, no. 12, pp. 2247-2253, Dec 2007.

[15] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local SVM approach," *International Conference on Pattern Recognition*, pp. 32-36, 2004.

[16] J. Wang, M. Zhu, Y. Zhao, *et al.*, "Time-involved Cutting Plane and Region Covariance Descriptor for Jump Action Recognition," *Chinese Conference on System Simulation Technology Application* Huangshan, China, 2011.

[17] L. Wang, T. Tan, H. Ning, *et al.*, "Silhouette analysis based gait recognition for human identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 25, no. 12, pp. 1505-1518, 2003.

[18] K. Mikolajczyk, T. Tuytelaars, C. Schmid, *et al.*, "A Comparison of Affine Region Detectors," *International Journal of Computer Vision,* vol. 65, no. 1-2, pp. 43-72, 2005.