Local Motion Analysis and Its Application in Video based Swimming Style Recognition

Xiaofeng Tong[1*], Lingyu Duan[2], Changsheng Xu[2], Qi Tian[2], Hanqing Lu[1]

[1]*National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing China 100080*
{xftong, luhq}@nlpr.ia.ac.cn

[2]*Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613*
{lingyu, xucs, tian}@i2r.a-star.edu.sg

Abstract

In this paper we study the problem of local motion analysis and apply it to swimming style recognition in broadcast sports video. Local motion analysis is challenging for two reasons: 1) local motion is usually buried in clutters involving complex motion from multiple objects; and 2) the process is more sensitive to noises compared to the recovery of global motion. However, an effective approach to local motion analysis is significant for understanding human activity from image sequences. In this work, we firstly extract the object-induced local motion by utilizing robust motion estimation and salient color. The object motion is accordingly characterized by compensated motion vectors and confidence measurement. Beyond a single image, we attempt to capture the motion periodicity over the local motion sequence. For each period, we locate a so-called salient frame within which we derive a compact representation to distinctly characterize an image sequence with repeated actions. Finally, we employ a hierarchical classifier to distinguish local motion based on periodicity and salient frames. Promising results have been achieved on swimming style recognition in broadcast sports video.

1. Introduction

Motion plays a critical role in video analysis. It is unique and intrinsic for characterizing the dynamics and evolution of an object or a scene along time.

Early work focused on global motion. A 2-, 2.5- or 3-D parametric transformation (e.g. 2-D affine, 2-D quadratic, 2-D projective, 2.5-D affine and 3-D affine) was usually used to model the transformation between two successive images. Recently, many statistical methods were introduced to represent motion patterns, such as directional slices [1], motion modes seeking [2], Gibbs model [3], Dirac and exponential model [4], ARMA process [5], motion histogram [6], and HMM based learning [7, 8], etc.

In terms of local motion, much work concerns human activity. Allmen *et al* [9] used a spatio-temporal (ST) feature curve recovered from a ST-cube to represent an image sequence of human motion. The ST-cube was also employed to detect periodic motion with motion templates [10]. In [11], motion recognition was conducted within the legs region of a pedestrian based on time-delay neural networks. A ROI-based motion analysis technique was utilized in [12] where a self-similarity matrix of an object was formed over time to show lattice-like patterns of interest to motion analysis. A descriptor was proposed in [13] to measure periodic motions in human sports activities.

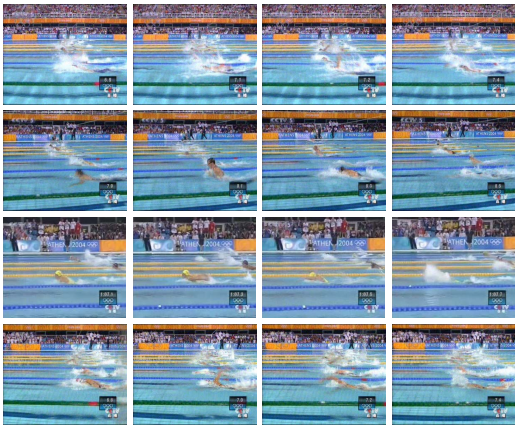

Figure 1. Image samples from swimming videos, back (1st row), breast (2nd row), butterfly (3rd row) and freestyle (4th row).

In this paper, we propose an approach for local motion analysis and apply it to video-based swimming style recognition. Local motion analysis is significant as the recovered global motion cannot discriminate between video clips of different swimming styles from the sports video indexing point of view. Examples are shown in Figure 1. The observed motion is produced by the mixture of camera motion and multiple objects motion. Moreover, the interesting local motion pattern is buried in cluttered envirements.

Human activity usually consists of periodically repeated actions. The motion behaviors within one period may compactly characterize what occurs in a lengthy sequence. Periodicity also gives the information about pace and speed. Hence our motion analysis is carried out at the level of motion periods. In addition, the robustness of motion classification can be improved by votes from multiple periods.

The frames within a period may differ in discriminating capabilities. As indicated in Figure 1, some frames have distinguishing features whereas some do not. Then we propose the concept of "salient frame", i.e., the frame with distinct features for effective classification.

Aiming to robustly recognize the swimming style, we propose a hierarchical classifier wherein domain knowledge is incorporated. Comparison study is conducted.

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00

2. Framework and methodology

The framework of our approach is illustrated in Figure 2. The player motion is represented by the features of motion vector values and the "outlier" confidence of local motion. A salient frame is located to exact compact and distinct motion features for motion recognition.

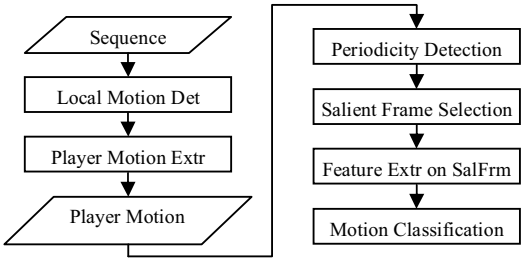

Figure 2. The overall framework

2.1 Local motion detection

The motion field is dealt with as two layers, namely, global motion (induced by a camera) and local motion (induced by objects). Local motion is considered as outliers during the regression for recovering global motion. A robust estimation technique *M-estimator*, which is less sensitive to outliers and is endued with a deterministic optimization scheme, is employed to robustly recover multiple motions [14]. The derived local motion is mapped to a grey-level image in which the intensity at each pixel is proportional to the confidence of associated local motion.

2.2. Player motion extraction

2.2.1. Player-color-mask (PCM) extraction

A PCM is to designate the player region with a distinguishing color within a region of interest (ROI). In our work PCM means the skin color region of a swimmer within the swimming pool. The extraction procedure is as below:

1) Dominant color detection

An accumulated color histogram is utilized to extract the dominant color over a sequence of frames, which models the water color of swimming pool as shown in Figure 3(a).

2) ROI segmentation

The dominant color mode is used to segment the ROI of swimming pool. Morphological operators, region filling and connection, and contour analysis are then applied to extract the complete ROI region as shown in Figure 3(b)..

3) Skin detection

The player region is located by identifying skin color. A simple uni-modal Gaussian with multi-variables is used to model skin colors. The detected skin region (white pixels) is shown in Figure 3(c).

4) PCM generation

The skin region within the ROI is regarded as PCM as shown in Figure 3(d).

2.2.2. Player-motion-mask (PMM) extraction

Local motion induced by the players within the ROI is taken as the player-motion-mask (PMM) as shown in Figure 3(e). In PMM, the intensity at each pixel corresponds to the confidence of local motion.

2.2.3. Player motion extraction

The player motion representation combines the compensated motion vector (magnitude and direction) and its confidence. In this work, the player motion is finally represented by the product of the motion magnitude and its confidence of local motion (the direction does not change). (See Figure 3(f)).

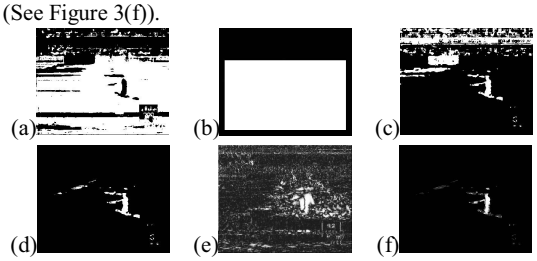

Figure 3. Player motion extraction. (a) dominant color (labeled by white); (b) playfield; (c) skin (white pixel); (d) player-color-mask; (e) local motion mask; (f) player-motion-mask;

2.3. Periodicity detection

Motion feature curves are computed from the image sequence by computing the mean squared magnitudes of the player motion frame by frame. We extract two feature curves $D_x(t)$ and $D_y(t)$ in horizontal and vertical directions.

$$D_x = \frac{1}{N}\sum_{i\in PCM} mv_{i,x}^2 \qquad D_y = \frac{1}{N}\sum_{i\in PCM} mv_{i,y}^2$$

where N is the number of pixels within the PMM, and $mv_i = \{mv_{ix}, mv_{iy}\}$ denotes the local motion vector at the i^{th} pixel.

We detect the periodicity in x- and y- direction, T_x and T_y, by using the autocorrelation of $\{D_i(.)\}$ and $\{D_i(.)\}$, respectively. For a periodic signal, the autocorrelation series must produce a uniform periodicity.

We search local maximums along an autocorrelation curve with a sliding window. Indices of local maximum are obtained. The final periodicity is estimated from the set of indices by the least square fitting.

2.4. Salient frame selection

A salient frame is defined according to the max-area motion region rather than the whole motion map. We firstly estimate the largest motion region, which is named as dominant local motion region (DLMR). DLMR is meant to deal with a motion map involving multiple local motion regions. Subsequently one salient frame containing DLMR is selected from a certain period in the image sequence.

As briefed below, four features have been defined for selecting salient frame in terms of the reliability, stabilization and prominence of a local motion region.

1) Reliability. It is defined as the average probability of local motion within the DLMR in a frame.

$$\mathrm{R} = \sum_{(x,y)\in DLMR} PMM(x,y) \Big/ N$$

where N is the total number of pixels in DLMR.

2) Stability. It is defined as the ratio of the mean to the variance of player motion probability within the DLMR.

$$S = \mu/\sigma, \quad \mu = \sum_{(x,y)\in DLMR} PMM(x,y) \Big/ N$$

$$\sigma = \sqrt{\sum_{(x,y)\in DLMR} [PMM(x,y) - \mu]^2 \Big/ N}$$

3) Magnitude. It is simply defined as the sum of player motion probability within the DLMR.

4) Eccentricity. $E = \max(h, w) / \min(h, w)$, where h and w are the height and the width of the min-bounding rectangle of the DLMR, respectively.

After normalizing four features over the whole sequence, we compute their product to evaluate the saliency. Finally, the frame with the highest value is selected as the salient frame within a period.

2.5. Feature extraction on a salient frame

Four shape features are defined over DLMR within salient frames for subsequent recognition. They are:

1) Orientation (O): shape central normalized moments: $O = \tan(2u_{11} / (u_{20}\text{-}u_{02})) / 2$.

2) Elongation (L): the ratio of the length to the width of the bounding rectangle.

3) Compactness (C): the ratio of the square perimeter to the area of a region.

4) Eccentricity (E): the same as the eccentricity defined for the salient frame selection above.

2.6. Motion classification

Based on the features of periodicity (T_x, T_y) and four salient frame features, (O, L, C, E), we design the classifier. Several comparison experiments are conducted as below: 1) by using periodicity only; 2) by using the concatenation of periodicity and salient frame features; 3) by using a hierarchical scheme supported by domain knowledge. The performance comparison is conducted between auto and manual salient frame selection. More details in Section 3.

3. Experimental results

The dataset comprises 95 video sequences listed in Table 1. One swimming style is involved for each sequence. All data comes from 2004 Olympic Games broadcast video.

TABLE 1. Experimental dataset

Type	back	breast	butterfly	freestyle
seq #	22	21	25	27
period #	188	107	137	210
frame #	3429	3288	3359	3414
time (sec)	137	132	134	137

3.1. Periodicity detection

The detected periodicity and the ground truth are compared in Figure 4. Due to twice arm swaying within one period for "back" and "freestyle", the detected periodicity is half of the real action periodicity. Hence, the speed of "back" and "freestyle" is faster than that of "breast" and "butterfly", although the decreasing order in terms of speed actually is "free", "butterfly", "back" and "breast". In order to evaluate the detection performance, we label the ground truth in the half real periodicity. The ground truth is decided to be the average of several periods involved in a sequence according to manual observation.

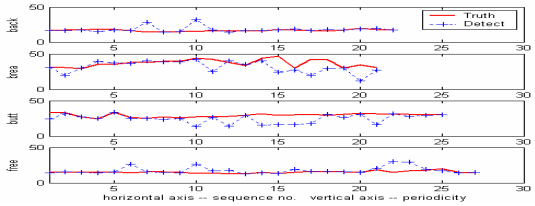

Figure 4. Periodicity detection

3.2. Recognition

A video clip often involves several action periods. Thus, more than one salient frames are generated for each clip. At the clip level, the recognition result is finally determined by the major voting of the recognition results of salient frames.

3.2.1. Hierarchical scheme

We propose a hierarchical classification scheme as shown in Figure 5. Domain knowledge is incorporated. At the first layer, we classify an incoming video clip into one of two major classes by using a k-NN classifier with the periodicity features. One class covers "back" and "freestyle" while the other class "breast" and "butterfly". At the second layer, two SVM classifiers are applied to discriminate the sub-classes. For SVM classifiers, a 4-dimensional feature vector on salient frames is applied.

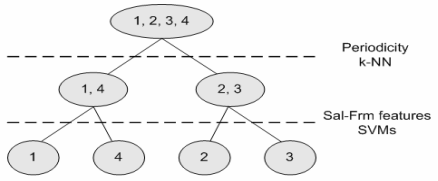

Figure 5. A hierarchical scheme for swimming style classification.

The classification performance at the top layer is listed in Table 2.

TABLE 2. The performance of periodicity-based classification

	(back, freestyle)	(breast, butterfly)
prec / recall	0.811 / 0.878	0.857 / 0.783

At the final recognition phase, we partition the dataset (training/testing) as follows: back (12 / 10), breast (11 / 10), butterfly (15 / 10), and freestyle (11/ 16). The performance is listed in Table 3.

TABLE 3. Recognition performance

	back	breast	butterfly	freestyle	prec	recall
back	10	0	0	0	0.67	1.0
breast	1	8	1	0	1.0	0.80
butter-	1	0	7	2	0.78	0.70
free-	3	0	2	11	0.79	0.69

As indicated in Table 3, the recognition performance of "freestyle" is the worst. The reasons are twofold. Firstly, "freestyle" is the fastest one. Fast motion speed may result in more noises and less accurate motion detection. Secondly, swimmer body parts are mostly buried by water and spray, which causes less accurate detection of PCM.

3.2.2. Hierarchy vs. Concatenation

We have compared the results in two cases: 1) by using only periodicity feature with k-NN; 2) by combining periodicity feature and salient frames related features with a SVM classifier. These results are further compared to that by the hierarchical scheme. See Table 4. A precision-recall (PR) curve is illustrated in Figure 6. The same training/testing dataset is applied to the above experiments.

It is clearly observed that a hierarchy scheme have yielded the best performance. The reason lies in the use of domain knowledge for the design of our classifier.

TABLE 4. The comparison of classification performance

prec / recall	back	breast	butterfly	freestyle
periodicity	0.50 / 0.36	1.0 / 0.48	0.56 / 0.72	0.49 / 0.67
concatenate	0.77 / 1.0	1.0 / 0.30	0.78 / 0.70	0.71 / 0.31
hierarchy	0.67 / 1.0	1.0 / 0.80	0.78 / 0.70	0.79 / 0.69

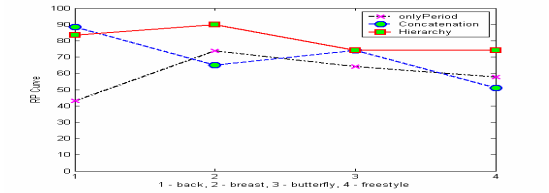

Figure 6. Classification performance comparison

3.2.3. Manu-SalFrm vs. Auto-SalFrm

To evaluate the effects of salient frame selection on the performance, we make three experiment cases as: Case 1 -- truth periodicity and manually selected salient frames (truth-Period + manu-SalFrm); Case 2 -- truth periodicity and automatically selected salient frames (truth-Period + auto-SalFrm); Case 3 – automatically detected periodicity and salient frame (auto-Period + auto-SalFrm). The results are listed in Table 5 and PR curves are shown in Figure 7.

TABLE 5. Performance comparison

prec / recall	back	breast	Butterfly	Freestyle
case 1	0.91/1.00	1.00/1.00	1.00/1.00	1.00/0.94
case 2	0.77/1.00	0.91/1.00	1.00/0.90	1.00/0.81
case 3	0.67/1.00	1.00/0.80	0.78/0.70	0.79/0.69

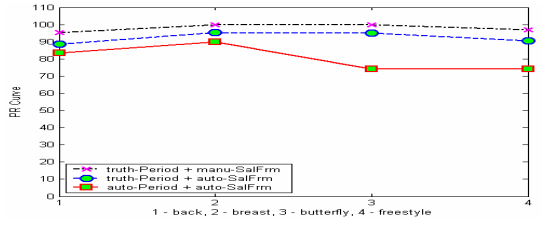

Figure 7. Curve for performance comparison

It is revealed that the selection of salient frames is feasible and effective for improving the performance of periodic motion recognition from an image sequence.

4. Conclusions

We have proposed an approach to local motion analysis for characterizing a video clip involving repeated actions. An application of video-based swimming style recognition was presented. The major contribution lie in the combination of local motion related parametric recovery and moment-based motion features to derive the concept of "salient frames" for distinguishing motion sequences. The selection of salient frames helps to yield a compact and effective descriptor for capturing periodic motions for video sequences. Future work includes the extension of period detection and salient frames selection to other sports video genres, e.g. Field & Track sports video.

5. Acknowledgement

This work is supported by The National Key Basic Research and Development Program (973) under Grant No. 2004CB318107, Natural Sciences Foundation of China under Grant No 60475010 and 60121302.

We would like to thank Prof. Michael J. Black at Brown University for providing his robust motion estimation code.

6. References

[1] Y.F. Ma, H.J. Zhang, "Motion Texture: A New Motion Based Video Representation", ICPR 2000, pp. 548-551.

[2] L.Y Duan, M. Xu, Q. Tian, and C. S. Xu, "Nonparametric motion model with application to camera motion pattern classification", Proc. of ACM Multimedia, 2004.

[3] R. Fablet, P. Bouthemy and P. Perez, "Nonparametric motion characterization using causal probabilistic models for video indexing and retrieval", IEEE Trans. on Image Processing, 11(4), 2002, pp.393-407.

[4] G. Piriou, P. Bounthemy, J. Yao, "Learned probabilistic image motion models for event detection in videos", ICPR 2004.

[5] S. Soatto, G. Doretto, and Y. Wu, "Dynamic textures", Proc. of ICCV 2001, pp. 439-446.

[6] A.K. Jain, A. Vailaya, and W. Xiong, "Query by video clip", Multimedia Systems, 7, 1999, pp.369-384.

[7] X.D. Sun, C.W. Chen, and B.S. Manjunath, "Probabilistic motion parameter models for human activity recognition", ICPR2002, pp.443-446.

[8] G.Xu, Y.F. Ma, H.J.Zhang, S.Q. Yang, "Motion Based Event Recognition Using HMM", ICPR 2002, pp. 831-834.

[9] M. Allmen and C.R.Dyer, "Cyclic motion detection using spatiotemporal surfaces and curves. ICPR1990, pp. 365-370.

[10] R. Polana and R.C.Nelson, "Detection and recognition of periodic, nonrigid motion", IJCV, 23(3), 1997, pp.262-282.

[11] B. Heisele and C.Wohler, "Motion-based recognition of pedestrians", ICPR1998, pp.1325-1330.

[12] R. Cutler and L. Davis, "Robust real-time periodic motion detection, analysis, and applications", IEEE Trans. on PAMI, 22(8), 2000, pp.781-796.

[13] F.X. Cheng, W. Christmas, and J. Kittler, "Periodic human motion description for sports video databases", ICPR 2004.

[14] M. Black, and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields", CVIU, 63(1), 1996, pp.75-104.