

Periodicity Detection of Local Motion

Xiaofeng Tong^{1*}, Lingyu Duan^{2,3}, Changsheng Xu², Qi Tian², Hanqing Lu¹, Jinjun Wang², Jesse S. Jin³

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, stuwj2}@i2r.a-star.edu.sg*

³*School of Design, Communication, and Information Technology, University of Newcastle, NSW 2308,
Australia,
jesse.jin@newcastle.edu.au*

Abstract

Periodicity is useful for compact representation of periodic motion and a reasonable selection of a proper temporal scale for periodic motion analysis. In this paper, we concern the periodicity detection of local motion within an interesting region and present an approach to automatically detect the motion periodicity inherent to local motion under complex condition. The task is challenging as local motion is usually buried in clutters with global motion and noises. Most existing methods have assumed a static camera and a labeled moving object region. We instead apply robust local motion estimation and an object localization method to extract the object motion. The object motion is characterized by the confidence based motion probability map and the motion vectors obtained by global motion compensation. The autocorrelation series of motion energy is then carried out to locate local maximum points. With the set of indices of local maximum points, we can estimate the basic periodicity through a least-square fitting. This method has been applied to swimming videos and got encouraging results.

1. Introduction

Periodicity is a significant descriptor for motion characterization. Within a lengthy motion sequence, the choice of a proper temporal scale and a compact motion representation is useful for periodic motion analysis.

Some previous work has studied periodic motion analysis in computer vision domain. Allmen *et al.* [1] utilized a spatio-temporal (ST) feature curve recovered from the ST-cube to represent an image sequence and human motion. The ST-cube was also employed to detect periodic motion based on motion templates [2]. In [3], the bounding box of a pedestrian's legs was used to perform motion recognition by a time-delay neural network. An area-based motion detection and analysis technique was introduced in [4], where a self-

similarity matrix showing a lattice-like pattern over time was used. In sports video domain, Cheng *et al.* [5] have proposed a template-based period-related motion descriptor and applied it to sports video classification. Unfortunately, the labeling of moving regions or the customization of sports template is difficult practically.

In this paper, we study the detection of local motion periodicity towards characterizing local motion and further indexing. The task is challenging as local motion is usually buried in clutters with the mixture of a global motion and noises. Our motion analysis is carried out within an interesting region instead of the whole motion field with an aim of the local motion based video indexing. Note that the motion analysis of an object of interest (OOI) does not require an accurate segmentation of a moving object, which is different from most existing methods. A robust local motion estimation technology is employed to identify the object region and estimate the optical flow field to extract local motion within a region of interest (ROI). With the local motion of ROI, we compute the motion energy based autocorrelation over a sequence along the horizontal and vertical directions, respectively. The local maximum points are located. With the indices of local maximum points, we finally estimate the motion periodicity by using a least-square fitting. This approach has been applied to swimming video analysis. For a broadcast swimming video, the recovered global motion (pan and zoom in general) is meaningless from a semantic video indexing point of view. Periodicity in swimming is an important characteristic. The detected periodicity can be further used for swimming style recognition and indexing [6], and drowning alarming [7].

The rest of this paper is organized as follows. The framework is introduced in Section 2. The detailed procedure is discussed in Section 3. Section 4 describes experiments and some applications. Conclusions are drawn in Section 5.

2. Framework

The flowchart of our approach is shown in Figure 1. It consists of two parts: object motion detection and motion statistics analysis based on periodicity estimation. Compared

* This work was performed when the author was visiting Institute for Infocomm Research, Singapore as an intern.

with most existing work on moving object extraction, our method only requires coarse object localization.

The framework includes: (1) Local motion detection via layered motion analysis and local-motion-mask (LMM) generation; (2) Object identification with an object-color-mask (OCM); (3) Object-motion-mask (OMM) generation with the combination of OCM and LMM; (4) Object motion extraction with LMM and motion vectors obtained by a global motion compensation; (5) Motion statistics obtained by the autocorrelation of object motion energy series; (6) Periodicity estimation over a set of index of local maximum points of autocorrelation series.

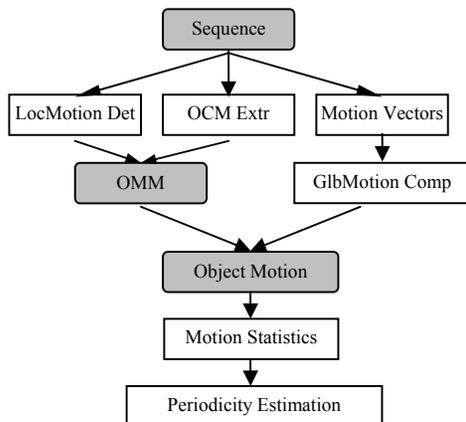


Figure 1. Flowchart of our proposed method

3. Methodology

3.1. Local motion extraction

We treat the motion field as two layers: camera induced global motion and object induced local motion. Local motion is considered as outliers during the regression for recovering global motion. A robust estimation technique M-estimator, which is less sensitive to an assumption violation and combined with a deterministic optimization scheme, is employed to accurately recover multiple motions [8]. The extracted local motion are mapped to an image in which each pixel's intensity is proportional to the confidence of local motion. The confidence can be derived by the regression. For motion analysis, this probabilistic description is more reasonable and robust than the hard decision that a pixel is absolutely treated as either local motion or global one.

3.2. Object-color-mask detection

To extract concerned local motion, we need detect moving object. For object localization or segmentation, color histogram back projection and boundary constraint is often used [9]. Accurate object segmentation is not only difficult but unnecessary in this work. In this paper, we only construct a coarse object-color-mask which designates the objects' region within the playfield in sports video. The procedure includes dominant color detection, playfield and object extrac-

tion as shown in Figure 2.

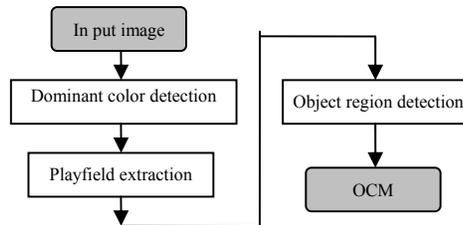


Figure 2. Procedure of OCM detection

1) Dominant color detection.

The playfield color is regarded as dominant color. We simply utilize an accumulated color histogram in HSV color space to extract the dominant color. Dominant colored pixels are displayed as white in Figure 3(a).

2) Playfield segmentation and extraction

The dominant color is used to segment playfield. Morphological operation, region filling and connection, and contour analysis are performed to ensure a complete playfield region (See Figure 3(b)).

3) Object region detection

In swimming video, a straightforward method of object region identification is to detect skin color. A simple and effective uni-modal Gaussian model with multi-variables is used. An example of skin color detection is shown in Figure 3(c).

4) Object-color-mask generation

The image that delineates skin regions within the playfield is regarded as an object-color-mask (See Figure 3(d)).

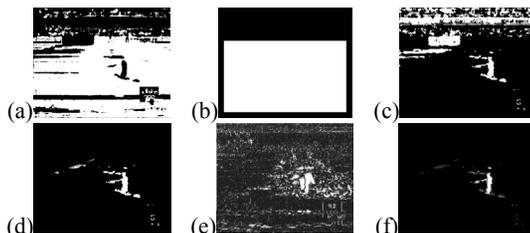


Figure 3. Object motion extraction in swimming video. (a) dominant color detection; (b) playfield; (c) skin region; (d) object-color-mask; (e) local-motion-mask; (f) object-motion-mask;

3.3. Object-motion-mask extraction

The local motion within an object-color-mask is taken as object-motion-mask, which designates not only the local moving region but also the confidence of local motion on each pixel. Confidence indicates the probability or reliability of a pixel in terms of its contribution to local motion. A sample OMM is shown in Figure 3(f).

3.4. Object motion

The object motion contains not only the local motion probability based confidence but also the remaining motion vectors (magnitude and direction) after global motion com-

penation. In this work, the object motion at each pixel is represented by the product of motion magnitude and its confidence (the direction does not change).

3.5. Motion statistics

Motion feature curves are computed from the image scenes by computing the mean squared motion magnitudes (motion energy) of the player motion frame by frame. In this work, two feature curves $D_x(t)$ and $D_y(t)$ are extracted in horizontal and vertical direction.

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

where N is the number of pixels within the OCM, and $mv_i = \{mv_{ix}, mv_{iy}\}$ represents the i^{th} pixel local motion generated by the original motion vector and motion confidence.

We detect the periodicity in x - and y - direction, T_x and T_y , by using the autocorrelation of $\{D_x(\cdot)\}$ and $\{D_y(\cdot)\}$ respectively (Figure 4). For a periodic signal, its autocorrelation series must be also periodic and has the same periodicity. The autocorrelation series of a signal of $\{s(\cdot)\}$ is defined as below:

$$\hat{R}_{ss}(\tau) = \frac{1}{N} \sum_{i=1}^N s(i) \cdot s(i + \tau)$$

where N is length of $\{s(\cdot)\}$.

Then we search local maximum points over autocorrelation series with a sliding window.

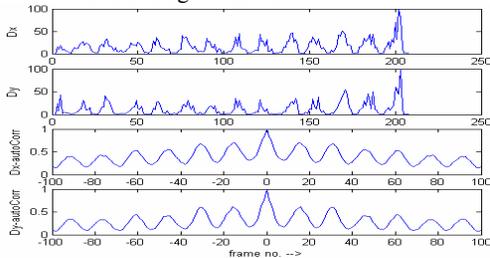


Figure 4. Motion energy and autocorrelation curves.

3.6. Periodicity estimation

The set of indices of local maximum points in auto-correlation series is denoted by $\{x_i, i = 0, 1, 2, \dots, N-1\}$. Generally, x_i is a multiple of the basic periodicity T_0 , i.e. $x_i = m * T_0$. But the coefficient m and the T_0 are unknown. In order to robustly estimate the T_0 , we perform operations as below.

- 1) Assume $y_i = \Delta x_i = x_{i+1} - x_i$, then we get a series $\{y_i | y_i = \Delta x_i = x_{i+1} - x_i, i \in [0, N-2]\}$. y_i also should be a multiple of T_0 .
- 2) Compute the medium of $\{y_i\}$. $T_c = \text{medium}\{y_i\}$. T_c is regarded as a candidate of the basic periodicity T_0 .
- 3) For each entry of $\{x_i\}$, we get a pair of parameters $\{k_i, x_i\}$, where $x_i = k_i * T_c$.

4) With the set of parameter pairs, we estimate the basic periodicity by using least square straight-line fitting. $\{k_i\}$ are treated as horizontal coordinates and $\{x_i\}$ corresponding vertical coordinates. The tangent of the slant angle θ of the fitted line is considered as an estimate of T_0 (See Figure 5).

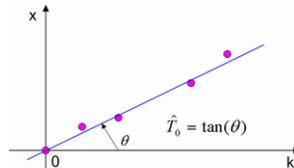


Figure 5. Periodicity estimation

4. Experimental results

We apply this approach to detect the periodicity for four swimming styles (i.e. back, breast, butterfly and freestyle). Generally speaking, there are three major shot types in broadcast swimming videos, i.e. long shot (above water), medium shot (above water) and underwater shot (See Figure 6). In our experiment, only medium shots are selected since medium shots can well capture the visual characteristics of swimming styles and have much more discrimination abilities. Those shots can be automatically selected by the shot boundary detection and shot classification.

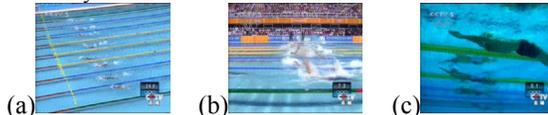


Figure 6. Three types of shots. (a) long shot; (b) medium shot; (c) underwater shot

Some sample images of swimming sequences within one period are shown in Figure 7. From these samples, we can notice the presence of the mixture of camera motion, player motion and noises (induced by audience, referee and spray). Intuitively, the period can be treated as an indicator of cadence and speed.

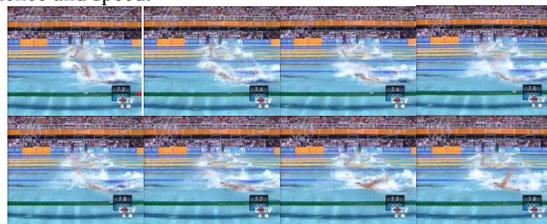


Figure 7. Sample images within a period

Table 1 lists the experimental data statistics including sequence, period, frame and time. The length of each sequence ranges from 44 to 314 frames (2 ~ 10 periods). Only one style is contained for each sequence. The video data are collected from Olympic Games 2004 broadcast video.

TABLE 1. Statistics of experimental data

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

4.1. Performance evaluation

We evaluate the performance by the mean and variance of an error between ground-truth and detected results.

$$e = |T_d - T_i| / T_i$$

where, T_d is the detected periodicity, and T_i is the ground truth.

4.2. Results

The performance figures are summarized in Table 2. The curve in Figure 7 illustrates the differences between the detected results and the ground truth. The ground truth is labeled as the average value of several periods within a sequence according to manual observation. The experimental results have revealed that the periodicities of “free” and “back” are smaller than that of “breast” and “butterfly”. This fact seems abnormal. It is because the action of arm sway happens two times within one period; the detected periodicity is half of the truth for “back” and “freestyle”.

For “back” and “free” styles, the falsely detected periodicities are often two times of the true values. It is probably due to the faster speed and the alternant arms protrusion above water. Fast moving speed results in much more spray, noise and inaccurate object identification with just color cue. But for “breast” and “butterfly”, the falsely detected results are mostly the half of the ground truth. The false alarms may arise from multiple players’ motion. In fact, what we have detected is the periodicity inherent to the whole local motion field without accurate moving object segmentation and tracking. That is, the major motion object with a dominant size directly affects the periodicity estimation.

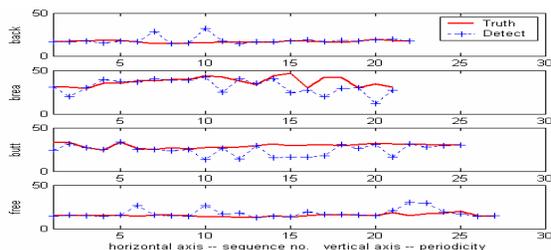


Figure 7. Periodicity detection result

TABLE 2. Error evaluation of periodicity detection performance

	back	breast	butterfly	freestyle
mean	0.1290	0.1689	0.1717	0.1973
var	0.2994	0.2006	0.2006	0.3249

4.3. Applications and discussion

As a significant descriptor of a periodic motion, periodicity can be further used for video analysis. In [6], we firstly

estimate the motion periodicity and further determine the salient frame within each period for recognizing swimming style. The average recognition performance for four swimming styles has reached up to 80.38% in terms of average precision and recall. Style analysis and recognition may be used for drowning alarm detection [7], where the motion speed, spray and activity were combined to predict possible drowning alarms.

The proposed approach could be extended to other sports, such as sprint and canoeing if an appropriate object identification scheme can be developed.

5. Conclusions

In this paper, we have proposed an approach for local motion analysis via periodicity detection under complex condition. This work is motivated by the difficulties of analyzing local motion in the presence of clutters consisting of global and multi-object motions. The contribution lies in the combination of the layered motion analysis and the autocorrelation of motion energy to estimate the basic motion periodicity. The idea can be widely used for local motion-based video classification.

6. Acknowledge

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

This work is supported by National Science Foundation of China (NSFC) under Grant 60475010 and 60135020.

7. References

- [1] M. Allmen and C.R. Dyer, “Cyclic motion detection using spatiotemporal surfaces and curves”, ICPR 1990, pp. 365-370.
- [2] R. Polana and R.C. Nelson, “Detection and recognition of periodic, nonrigid motion”, IJCV, 23(3), 1997, pp.262-282.
- [3] B. Heisele and C. Wohler, “Motion-based recognition of pedestrians”, ICPR 1998, pp. 1325-1330.
- [4] R. Cutler and L. Davis, “Robust real-time periodic motion detection, analysis, and application”, IEEE Trans. on PAMI, 22(8), 2000, pp.781-796.
- [5] F.X. Cheng, W. Christmas, and J. Kittler, “Periodic human motion description for sports video database”, ICPR 2004.
- [6] X. Tong, L. Duan, C. Xu, Q. Tian, and H. Lu, “Local motion analysis and its application in video based swimming style recognition”, submitted to BMVC 2005.
- [7] H. Eng, K. Toh, A. Kam, J. Wang, and W. Yau, “An automatic drowning detection surveillance system for challenging outdoor pool environments”, ICCV 2003, pp.532-539.
- [8] M. Black and P. Anandan, “The robust estimation of multiple motions: parametric and piecewise-smooth flow fields”, CVIU, 63(1), 1996, pp.75-104.
- [9] M. Swain, and D. Ballard, “Color indexing”, IJCV 1991, pp. 11-32.