Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 6–21
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Generic temporal segmentation of cyclic human motion

A. Branzan Albu[a], R. Bergevin[b,*], S. Quirion[b]

[a]*Department of Electrical and Computer Engineering, University of Victoria, Victoria, Canada*
[b]*Department of Electrical and Computer Engineering, Laval University, Que. City, Canada*

Received 17 May 2006; received in revised form 15 February 2007; accepted 13 March 2007

Abstract

A method is proposed for the temporal segmentation of cyclic human motion from video sequences. The proposed method is divided into three processing steps. Once silhouettes and body part locations are obtained, a set of individual 1-D signals representing motion trajectories of body parts is extracted for the entire sequence. The second step performs the individual segmentation of all signals in the set in order to localize their periodic segments. In the final step, all individual segmentations are coherently merged into a global segmentation for the entire sequence and set of signals. The proposed approach has been successfully tested on a variety of sequences containing cyclic activities such as aerobic exercises and walking along different directions.

Keywords: Human motion analysis; Periodicity analysis; Temporal segmentation

1. Introduction

Human motion analysis is a very active topic in computer vision. Research in this field is driven by theoretical challenges specific to video understanding, as well as by the wide spectrum of applications in surveillance, perceptual interface design, and health. One may identify two main themes in video-based human motion analysis, related to: (a) biometrics, namely gait-based person identification and (b) activity recognition. Since walking is an activity per se, these two goals can be reformulated as: (a) recognizing a person from the way he performs an activity and (b) recognizing the activity itself.

The recognition problem can be tackled in both cases only after an accurate detection of the temporal boundaries of the activity of interest. However, most of the published work in activity recognition, such as Polana and Nelson [1] and Ben-Arie et al. [2], does not address the boundaries detection problem as each chosen experimental video sequence contains a single activity instance. This choice allows one to focus on finding the most appropriate motion representation for activity recognition purposes. However, it is unclear whether the generation of an activity-specific motion representation would function well for the detection of that activity in sequences containing multiple activities. The missing link to be addressed in this paper is the temporal segmentation of activities prior to activity representation, analysis and recognition.

Though highly desirable, a generic segmentation method is not easily attainable as no clear definition exists of what generic pattern of motion may represent a human activity. For instance, Rui and Anandan propose in Ref. [3] an approach for temporal segmentation based on the temporal discontinuities of the spatial pattern of image motion that captures the action. Their approach results in a fine-grained segmentation with segments corresponding to simple, continuous motions, such as an unidirectional arm swing. Such segments would have to be further aggregated in order to form a higher level description of a human action. One may conclude that defining a human action as a temporally consistent motion results in temporal oversegmentation.

Gao et al. propose in Ref. [4] a method for the temporal segmentation of activities in a dining room. Their work is also based on the concept of temporal consistency of human actions and involves mainly hand-head relative motion analysis of seated subjects. Hence, adding contextual constraints and focusing on a specific type of human action eliminates the oversegmentation problem.

* Corresponding author. Tel.: +1 418 656 2131x5173; fax: +1 418 656 3159.
E-mail addresses: aalbu@ece.uvic.ca (A.B. Albu), bergevin@gel.ulaval.ca (R. Bergevin), squirion@gel.ulaval.ca (S. Quirion).

0031-3203/$30.00
doi:10.1016/j.patcog.2007.03.013

Min and Kasturi describe in Ref. [5] a method for the high-level segmentation of human actions which uses multiple motion trajectories of body parts. The motion trajectories are first extracted by locating significant motion points and a color-optical flow-based tracker. Next, motion trajectories are used as features for the temporal segmentation of human activities. The human activities of interest are ballet steps, thus defined on a semantic level rather than from a spatiotemporal consistency perspective. A priori knowledge about the activities of interest is embedded in the training phase of the temporal segmentation, which involves HMM models for hands and legs trajectories.

The approach proposed in this paper focuses on the temporal segmentation of cyclic activities, a significant subset of human activities. According to Ref. [1], cyclic activities are those composed of regularly repeating sequences of motion events. Locomotion-related human activities, such as walking and running, are present in surveillance and medical monitoring contexts and they are cyclic in nature. Other examples of human activities that are cyclic in specific contexts include eating, reading, writing, playing, physical training, dancing, bicycling, clapping, swimming, working, etc. The cyclic nature of activities of interest allows for formulating their definition as a trade-off between low-level temporal consistency and high-level semantic definitions. In the context of our work, the motion events are described as multiple trajectories of body parts extracted using skeletal topologies.

Based on the above definition, we propose a new generic method for the temporal segmentation of cyclic human activities from a video sequence. The main idea behind our approach consists in relating the change in the human activity to discontinuities in the periodicity of the signals representing the activity. For instance, a change in the walking direction is interpreted as a separator between two walking activities, since this change is reflected as a temporal break between the two corresponding sets of periodic signals. Moreover, our approach is able to differentiate activities in terms of the composition of their set of periodic signals, such as walking followed by simultaneous walking and waving one hand.

Our main contribution lies in the capability of the proposed approach to accurately detect temporal boundaries of cyclic activities without using activity-specific prior knowledge, activity modeling, or training. In fact, the only prior knowledge embedded in the proposed approach is the cyclic character of the activities of interest. The experimental results are to show that the proposed approach provides reliable results not only on sequences containing ample body motions, such as aerobic exercises, but also on sequences involving more common motions such as human walking. Preliminary results of our study appeared in Quirion et al. [6]. The present paper contains a comprehensive description of the proposed method, which features significant conceptual updates with respect to Ref. [6] and is extensively validated using new performance evaluation measures over an enriched experimental database.

The proposed method is divided into three processing steps. Once silhouettes and body part locations are obtained, a set of individual 1-D signals representing motion trajectories of body parts is extracted for the entire sequence. The second step performs the individual segmentation of all signals in the set in order to localize their periodic segments. In the final step, all individual segmentations are coherently merged into a single global segmentation for the entire sequence and set of signals. The rest of the paper is structured as follows. Section 2 presents related work in the field of periodic motion analysis. The detailed description of the proposed method is given in Section 3. Section 4 presents the results of an extensive experimental validation. Section 5 draws conclusions and describes future work.

2. Related work in periodic motion analysis

Periodic motion instances, often a direct manifestation of basic rhythms of life, are to be found in the natural world. This makes periodicity a powerful cue for extracting information about topics ranging from marine life [7] to animal and human gait [8–10] and to human gestures analysis [11].

The literature on video-based periodic motion analysis is structured along a few major research directions. The two main ones are detailed below and supported by appropriate references which were selected among the most relevant field-specific contributions in the last decade.

Periodicity can be used for discriminating between human and non-human motion, and thus for detecting pedestrians in a surveillance context. Cutler and Davis [9] differentiate between periodic (human), periodic (animal) and aperiodic (translational) motion by computing an inter-frame similarity matrix and its normalized autocorrelation for each type of motion. They extract information about the period of motion by fitting a lattice on the autocorrelation matrix, a technique inspired from earlier work on spatial periodic texture analysis [12]. Ran et al. [13] describe a method for detecting pedestrians in videos acquired from moving cameras. Their method is based on the extraction of a periodic pattern for each walking pedestrian by using a twin-pendulum model. A similar idea is used in Ref. [14] for classifying objects (pedestrians, cars) from infrared videos by analyzing the periodic signature of their motion pattern with finite frequencies probing.

Periodicity also plays a major role in approaches for gait-based person identification, where gait is described by pixel- or region-based oscillations. For example, Little and Boyd [15] use the discrete Fourier transform to first extract the fundamental frequency of gait, and then to measure relative phase differences between motion signals computed from optical flow. They conclude that some phase features are consistent for one person, and show significant statistical variation between persons. Tsai et al. [16] detect gait cycles using autocorrelation and Fourier transform of the smoothed spatio-temporal trajectories of specific points on the walking human body. They found that cyclic motion is helpful in reducing the overhead of the motion-based recognition by performing cycle segmentation as a preprocessing step. Cunado et al. [17] use periodicity information in representing the periodic hip rotation during walking by Fourier series. They use this representation in conjunction with velocity Hough transform for building a feature-based, subject-representative gait model.

While not questioning the merits of the above-mentioned work in periodic motion analysis, one may notice a general limitation in the applicability of the existing techniques. All methods are based on the assumption that periodic motion occurs continuously, i.e. people walk in a regular way, without stopping or changing their activity pattern. This assumption is not valid in real-life situations. Periodic motion (i.e. gait) is usually interrupted by stops, changes in the walking direction, or other aperiodic, human activities. This is why an accurate temporal segmentation of periodic human activities from video data is necessary prior to periodic motion analysis.

Yazdi et al. [18] describe a temporal segmentation method for cyclic activities using a 2-D inter-frame silhouette-based similarity plot. However, their analysis applies only to symmetrical cyclic activities, where the motion performed during the first semi-cycle is repeated in the opposite direction during the second semi-cycle. Another limitation is that all cycles must be complete, which is not to be the case in the proposed method.

This paper proposes a new method for the temporal segmentation of cyclic activities from a set of 1-D signals corresponding to the spatiotemporal trajectories of body parts. Experimental results are to show that our method is able to accurately detect temporal boundaries of cyclic activities in video sequences containing multiple activities. The detailed description of the proposed method is to be found in the following section.

3. Proposed approach

The proposed approach describes human motion in terms of a set of 1-D signals associated with the spatiotemporal trajectories of a limited number of feature points located on the human body. One spatiotemporal trajectory can be described by one or more 1-D signals in the set. A cyclic action involving one or more body parts will translate into a periodic segment on at least one signal in the set.

3.1. Signal extraction

Signal extraction is a preprocessing step which must first deal with the detection of significant points; second, it has to describe the trajectory of each significant point with a number of 1-D signals. The proposed work has used two different methods for detecting significant points which will be briefly detailed below. The generation of the signal set following each method of significant point detection will also be explained. It is worth mentioning that the proposed segmentation approach is compatible with any other method of signal extraction, provided that this method successfully converts a cyclic activity into a set of signals containing a subset of periodic segments.

3.1.1. Detection of significant points by skeleton fitting

A sequence of binary silhouettes is first obtained from each input sequence via a simple differential background subtraction technique. Next, a 14-segment skeleton is fitted to each silhouette using the method proposed by Vignola et al. [19]. This first method for significant point detection (thereafter called "SPD1") performs a sequential skeleton fitting process on a frame-by-frame basis, as shown in Fig. 1; the edges of the skeleton represent the significant points to be detected. A six-segment torso model is first fitted to the silhouette by using information from the distance transform (DT). Specifically, the brightest points of the DT image form a medial axis of the human silhouette which can be viewed as a rough partial estimate of the skeleton. Next, the configuration of the skeleton is completed with an iterative algorithm searching for local maxima in the DT image of the silhouette (see Fig. 1f).

A limitation of detecting significant points by a two-dimensional skeleton fitting method is the sensitivity to the pose of the subject. It was found in our experiments that robust and reliable results are obtained for the frontal pose only.

3.1.2. Detection of significant points by motion tracking

A second method for significant point detection (thereafter called "SPD2") is the one proposed by Jean et al. [20]. It is used here to automatically detect and track six significant points (the centers of mass of the head, the hands, the feet, and the entire silhouette). Tracking is fully automatic, with no manual initialization required. Feet are detected in each frame by first finding the space between the legs in the human silhouette. The issue of feet self-occlusion is handled using optical flow and motion correspondence. Skin color segmentation is used to find hands in each frame and tracking is achieved by using a bounding box overlap algorithm. The head is defined as the center of mass of a region filling a predefined percent in the upper silhouette. Fig. 2 shows a typical result of significant point detection in a walking sequence. The detection of significant points from motion tracking yields robust results regardless of the pose of the subject.

3.1.3. Generation of the signal set

The set of 1-D signals is used to describe the spatiotemporal trajectories of the detected significant points. Since periodic segments on these 1-D signals must correspond to cyclic activities, it is required to discriminate between common and relative motion. Indeed, some cyclic human activities (e.g. walking) exhibit common translational motion. In order to minimize the impact of common motion on the temporal segmentation of cyclic activities, information about relative and common motion will be explicitly stored in different signals.

When detecting significant points with SPD1, relative motion is described using the temporal variation of angles at joints of adjacent segments, as well as the spatiotemporal trajectory of relative x and y positions with respect to the adjacent joint closest to the torso (see Fig. 3 for an example). When detecting significant points with SPD2, relative x and y positions of the points corresponding to head, feet, and hands are computed with respect to the silhouette's center of mass. Angles between all pairs of segments defined by the silhouette's center of mass and a significant point were also tracked over time.

Torso motion can be considered as an accurate approximation of common motion when using SPD1. Therefore,

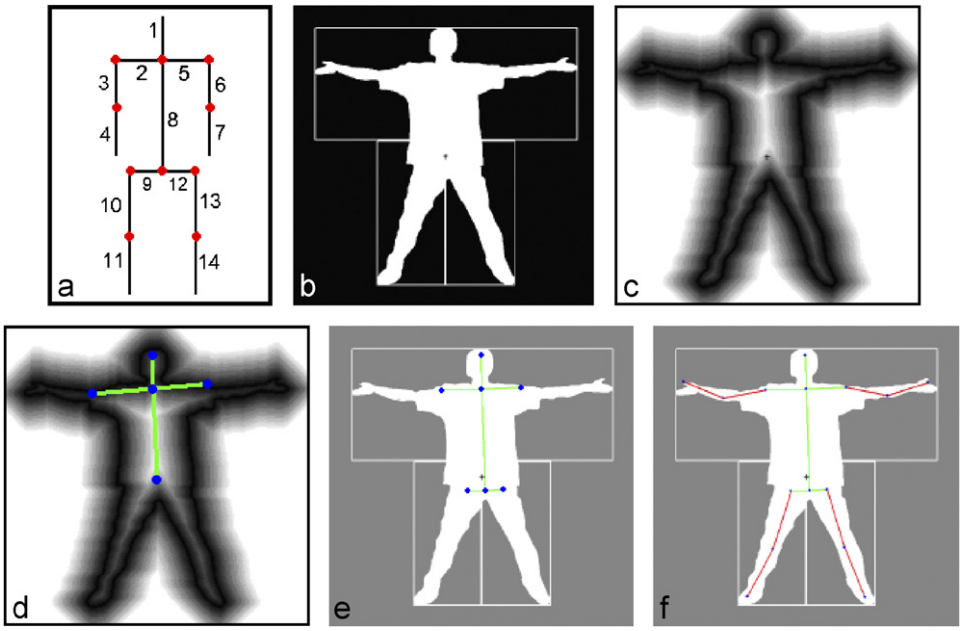

Fig. 1. (a) Articulated 14-segment skeleton, with annotated indexes of each segment; (b) binary silhouette resulting from background subtraction, and its division into four rectangular search boxes for further skeleton fitting; (c) DT of the silhouette in b; (d) torso fitting along the medial axis of the DT; (e) six segment torso model superimposed onto the binary silhouette; and (f) final result obtained after the sequential fitting of all segments corresponding to arms and legs.

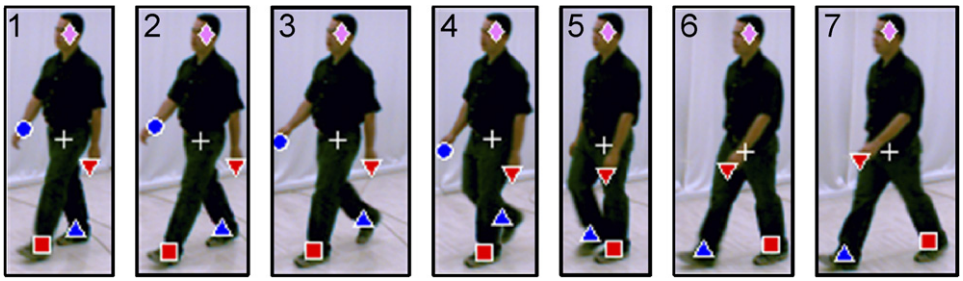

Fig. 2. Detection of six significant points during human walk.

common motion is described in this case by the spatiotemporal trajectories of the x and y coordinates of the torso segment, as well as by the temporal variation of the angle between the torso and the vertical image axis (see Fig. 4). When extracting significant points with SPD2, common motion is roughly approximated by the motion of the silhouette's center of mass; therefore, it is described through the spatiotemporal trajectory of the x and y coordinates of the silhouette's center of mass.

The temporal variations of angles and of x and y locations for all significant points are stored into a set of 1-D signals describing the activity content of the analyzed video sequence. The number of signals in the set is 34 (11 x-trajectories; 11 y-trajectories; 12 angles) when working with SPD1 and 22 (6 x-trajectories; 6 y-trajectories; 10 angles) when working with SPD2. Individual 1-D signals depict local translational and rotational motion of body parts occurring during human actions; these motions are strongly inter-related and constrained. Though it is possible to consider anatomical constraints for combining information from individual signals, initial attempts at doing so were not conclusive. Besides, a signal weighting scheme would be appropriate if the study of a specific motion (i.e. upper body motion) was targeted. This is not the case here since the proposed approach addresses the detection of generic cyclic motion. Therefore, experimental results in this paper are obtained without imposing anatomical constraints, nor using a weighting system for the signals in the set. These results indicate that the mere redundancy of periodic features found on more than one signal in the set is usually a sufficient cue for the segmentation of cyclic activities when no a priori knowledge of expected motions is available.

3.1.4. Discussion

Working with two different methods for significant point detection allowed the robustness of the segmentation algorithms to be explored with respect to noise generated by signal

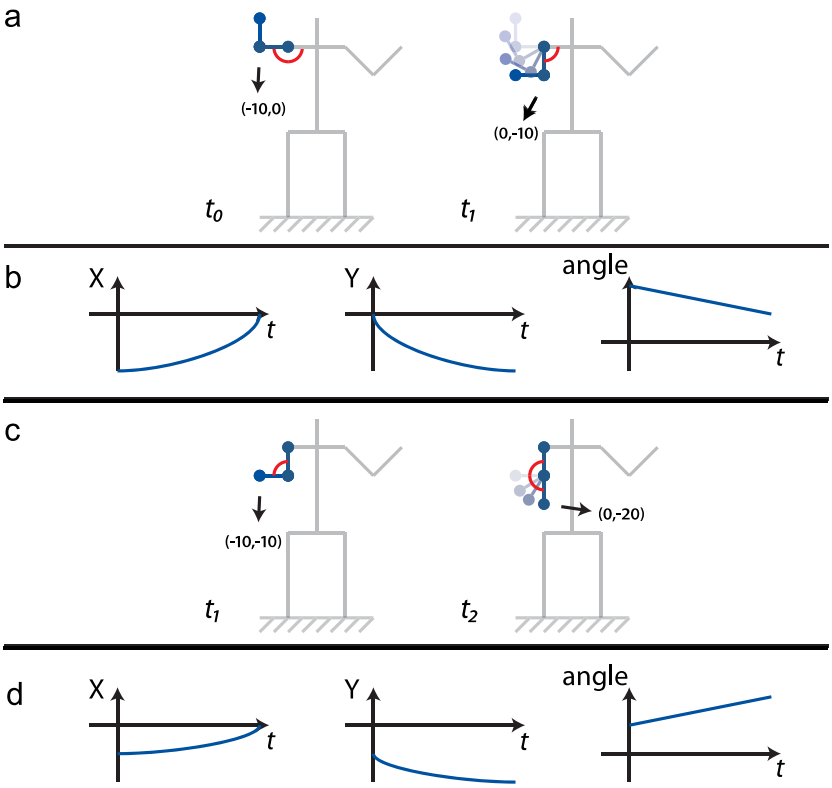

Fig. 3. (a) Relative motion of the significant point corresponding to a shoulder joint; (b) describing motion in (a) with a temporal plot of angle and spatial coordinates; (c) relative motion of the significant point corresponding to an elbow joint; and (d) describing motion in (c) with a temporal plot of angle and spatial coordinates.

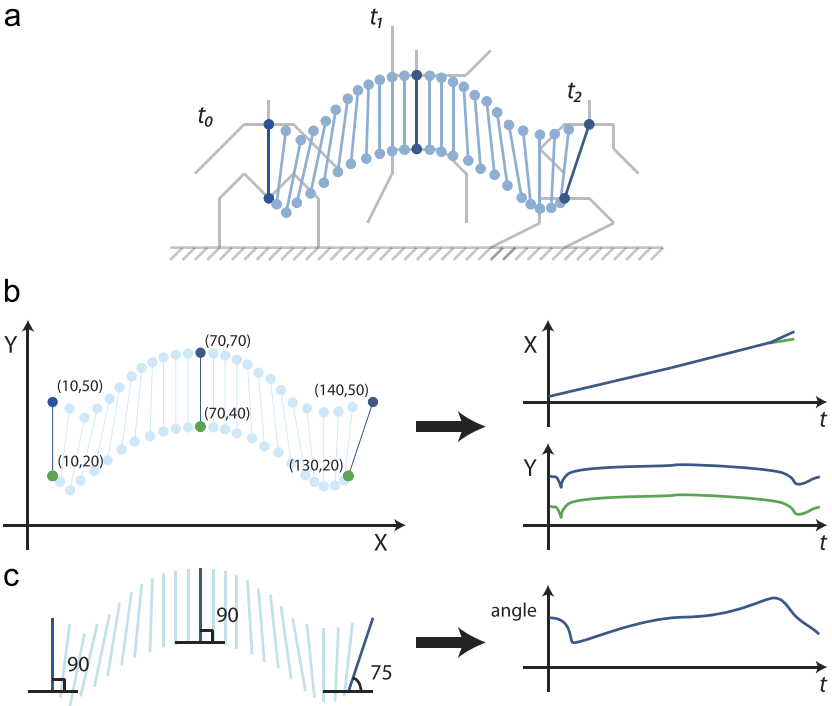

Fig. 4. Description of common motion: (a) temporal variation of the torso orientation when jumping to the right; (b) temporal plot of X and Y positions of torso joints; and (c) temporal plot of the angle between the torso and the horizontal axis.

extraction. For instance, SPD1 works on a frame-by-frame basis and it does not impose any temporal smoothness constraint on silhouettes in consecutive frames. As a result, the skeleton sequences contain "jitter" noise.

While generating smoother spatiotemporal trajectories for the tracked body parts, SPD2 extracts a smaller number of feature points. This enabled an exploration of the impact of the low redundancy of the periodic information contained in the signal set on the final result of cyclic activity segmentation. Moreover, angle signals extracted with SPD2 do not describe the articulated motion of anatomic joints; working with such signals enabled an investigation of the adequacy of a low-level, non-anatomical model for the detection of articulated human periodic motion.

3.2. Individual signal segmentation

It can be shown that for a discrete signal S composed of l_s samples (frames), there are 2^{l_s-1} possible segmentations or partitions, where each segment of the partition corresponds to a given number of consecutive frames. Even by adding constraints such as a minimum number of frames for each segment, the number of possible segmentations remains exponential in l_s. Hence, a brute force approach to detect the temporal boundaries of cyclic activities in a sequence is not appropriate. Instead, a deterministic greedy algorithm is proposed where the segments of the partition are extracted sequentially, beginning with the most periodic segments. In order to rank signals according to their periodicity, a new periodicity score is proposed.

3.2.1. Periodicity score

The decision whether a signal is periodic or not is binary. However, our approach does not focus on the detection of pure periodic signals since signals extracted from cyclic body motion usually exhibit noise, local irregularities and slight variations in amplitude/period. Non-ideal periodic signals have also been studied by Seitz and Dyer [21], who introduced the notion of period trace. However, their approach deals with the quantification of local irregularities, as well as with recovering the mean rate of increase/decrease of the period. Such measurements are not applicable in the context of our work.

The proposed approach for assessing the periodicity of 1-D signals is based on autocorrelation and thus similar to some extent to the method proposed by Cutler and Davis [9]. The main idea behind our approach is to compare the autocorrelation of a non-ideal periodic signal of average period c_S with the one of an ideal periodic signal of exact period c_S; more specifically, the corresponding maxima of the two autocorrelation functions will be compared. Let us consider two ordered sets of indexes of autocorrelation maxima: M_S for the non-ideal periodic signal and E_S for the periodic signal. E_S can be expressed as $E_S = (0, c_S, 2c_S, \ldots, nc_S)$ where n is the number of cycles included in the periodic signal. It is assumed that the cardinal of the two sets is identical, $|M_S| = |E_S|$ and therefore a biunivocal correspondence exists between the two sets.

The periodicity score Ψ_S is designed as a measure of proximity between pairs of corresponding maxima. For each pair, Ψ_S depends on their difference in lag normalized by the cycle length:

$$\Psi_S \propto 1 - \frac{|E_S(i) - M_S(i)|}{c_S}. \tag{1}$$

Ψ_S also depends on the difference in magnitude of the autocorrelation function A_S. For the ideal periodic signal, $A_S(E_S(i)) = 1$ for all $i = 1 \ldots n$. Therefore,

$$\Psi_S \propto A_S(M_S(i)). \tag{2}$$

The final expression of the periodicity score is obtained via averaging over the entire set of pairs of maxima:

$$\Psi_S = \frac{1}{|M_S| - 1} \sum_{i=2}^{|M_S|} \left(1 - \frac{|E_S(i) - M_S(i)|}{c_S}\right) \cdot A_S(M_S(i)). \tag{3}$$

The score of a periodic signal is equal to one and it decreases as the signal becomes less and less periodic. The score may be negative for degenerate cases (i.e. difference in lag greater than c_S) although such cases were never encountered in experiments.

To eliminate multiple partial detections of the same periodic segment, long periodic segments are preferred. Length has to be favored in periodic segments only, and therefore a threshold η_l is needed to distinguish between what is considered periodic and what is not. This threshold is used for defining a length-normalized periodicity score as follows:

$$\mathbf{Y}_{S_{[i,j]}} = \eta_l^{1-(j-i+1/l_S)} \cdot \Psi_{S_{[i,j]}}^{(j-i+1/l_S)}. \tag{4}$$

As the length of the segment $[i, j]$ approaches l_S, $\mathbf{Y}_{S_{[i,j]}}$ approaches $\Psi_{S_{[i,j]}}$. Also, as the length of segment $[i, j]$ approaches 0, $\mathbf{Y}_{S_{[i,j]}}$ approaches the threshold η_l. In other words, length improves the score of a periodic segment (i.e. a segment $[i, j]$ with $\Psi_{S_{[i,j]}} > \eta_l$) but decreases the score of a non-periodic segment.

3.2.2. Greedy segmentation

The proposed segmentation algorithm works iteratively. It first extracts the 'best' (most periodic) segment in the signal by using a simple global maximum search algorithm (see Algorithm 2). This segment is included in the segmentation set provided that its length surpasses a minimum length β and its periodicity score is above a threshold η_h. The remaining portions of the signal are processed in the same fashion until no segments satisfying the length and periodicity criteria are to be found. The pseudo codes for the greedy segmentation, as well as for the extraction of the best segment are given by Algorithms 1 and 2, respectively.

Algorithm 1. $\mathbb{SEG} = GreedySegmentation(S, \beta, \eta_h)$

(1) initialize segmentation set $\mathbb{SEG} \leftarrow \emptyset$
(2) define set of segment search spaces $\mathbb{X} \leftarrow \{(1, length(S))\}$
(3) WHILE $\mathbb{X} \neq \emptyset$
  (a) pick at random (I, J) from $\mathbb{X}$
  (b) IF $J - I + 1 > \beta$
    (i) $(i, j) \leftarrow BestSegment\ (S_{[I,J]})$
    (ii) IF $\Psi_{S_{[i,j]}} > \eta_h$
      (A) update segmentation result $\mathbb{SEG} \leftarrow \mathbb{SEG} \cup \{(i, j)\}$
    (iii) END IF
    (iv) update set of search spaces $\mathbb{X} \leftarrow \mathbb{X} \cup \{(I, i), (j, J)\}$
  (c) END IF
  (d) update set of search spaces $\mathbb{X} \leftarrow \mathbb{X} - \{(I, J)\}$
(4) END WHILE

Algorithm 2. $(i, j) = BestSegment(S)$

(1) initialize $(i, j) \leftarrow (0, 0)$
(2) FOR $m \leftarrow 1$ to $length(S)$
  (a) FOR $n \leftarrow m + \beta - 1$ to $length(S)$
    (i) IF $\mathbf{Y}_{S_{[m,n]}} > \mathbf{Y}_{S_{[i,j]}}$
      (A) $(i, j) \leftarrow (m, n)$
    (ii) END IF
  (b) END FOR
(3) END FOR

Algorithm 1 returns a set of periodic segments belonging to the same signal S, while algorithm 2 returns the temporal boundaries $[i, j]$ of the most periodic segment in S. The segment search spaces used in Algorithm 1 are contiguous parts of the signal defined by their minimum and maximum indexes; they are used for limiting the search of the best segment to a specific part of the signal.

Fig. 5 presents a complete matrix of values for the length-normalized score as computed in order to obtain the 'best' segment in terms of length-normalized periodicity score.

3.3. Global segmentation

The aim of the global segmentation step is to detect the temporal boundaries of cyclic human activities manifested as periodic segments on at least one individual signal in the signal set. Cyclic human activities typically give rise to a set of partially overlapping periodic segments located on different signals. To extract the precise location of temporal boundaries for each cyclic activity, the proposed approach uses a global periodicity score and a greedy algorithm for combining individual signal segmentations.

3.3.1. Global periodicity score

For one segment defined by its temporal boundaries $[i, j]$ (with $i < j$) the global periodicity score is computed over the entire set of 1-D signals extracted from the initial video sequence. This score measures to what extent the segment isolates a periodic portion of the signal set. It is computed as a sum of the corresponding individual periodicity scores which are above the threshold η_l. This threshold is less strict than the one used in the individual segmentation ($\eta_l < \eta_h$) as false detections have already been addressed using the high threshold during individual signal segmentation. However, the η_l threshold is needed to insure that non-periodic segments with low individual scores do not sum up to a significant global score. Formally, the global periodicity score is expressed as

$$\mathcal{G}[i, j] = \sum_{k=1}^{n} \Psi^*_{S_{k[i,j]}}, \tag{5}$$

where the summed elements are

$$\Psi^*_{S_{k[i,j]}} = \begin{cases} \Psi_{S_{k[i,j]}} & \text{if } \Psi_{S_{k[i,j]}} > \eta_l, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Formulating the global periodicity score as a sum of individual scores above a threshold enables an exploitation of the redundancy of the periodicity information contained in the signal set. A segment $[i, j]$ corresponding to periodic portions on several individual signals in the set is likely to represent a cyclic human activity and thus receives a high global periodicity score. However, localized periodic motions which translate into few individual periodic segments are not disfavored due to the design of the greedy approach for combined segmentation. The need for normalizing the global periodicity score through averaging is not justified by the further use of this score; moreover, averaging may negatively impact the extraction of a localized periodic motion described by a small number of strong periodic segments.

3.3.2. Greedy algorithm for combining individual segmentations

Since a cyclic human activity is represented by at least one periodic segment located on one individual signal, each periodic segment detected in the individual segmentation step is an input candidate for the global or combined segmentation. Hence, the global periodicity score is computed for each candidate using Eq. (5). The result of the combined segmentation is the highest scoring non-overlapping subset of candidates.

Given the high number of possible combinations, a greedy combination algorithm is used where the best candidate, according to the global periodicity score, is identified and retained at each step. A straightforward solution consists in iteratively finding the highest scoring candidate from the current set of candidates, and adding it to the segmentation set before updating the set of candidates accordingly. The simplest update would

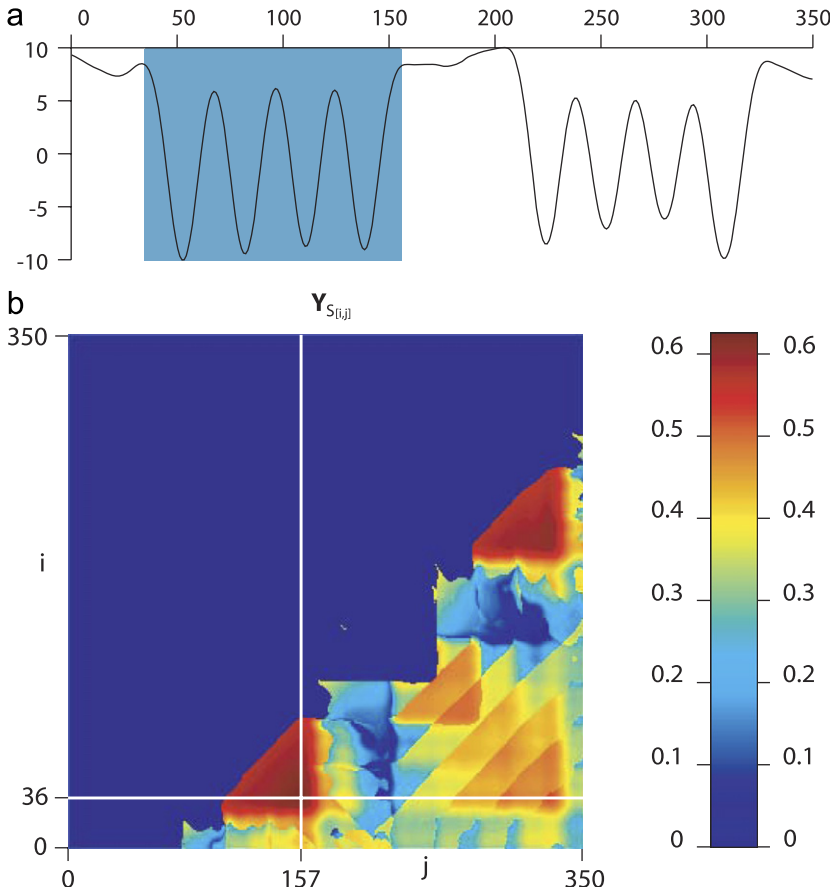

Fig. 5. (a) One-dimensional motion signal. Highlighted window represents a periodic segment of the greedy segmentation. (b) Length-normalized periodicity scores matrix. The score corresponding to the periodic segment in (a) is indicated by intersecting lines.

consist in removing the chosen candidate along with all partially overlapping candidate segments in the set. However, this approach has resulted in many missed detections in sequences of cyclic actions with close temporal boundaries. To increase the robustness of the global segmentation, the update discards only the overlapping portions of the remaining candidates. The remaining parts, called difference segments, are tested for periodicity and length. If their individual periodicity score exceeds η_h on at least one signal, then they are consistent with the set of candidates and therefore included in it. The pseudo code for combined segmentation is given by Algorithm 3.

4. Experimental results

The proposed approach aims at the temporal segmentation of generic cyclic activities from video sequences. Therefore, the experimental database needs to be carefully assembled in order to enable a comprehensive validation. The content of this section is structured as follows. The design of the experiment is described in Section 4.1, while the results of the quantitative performance analysis are presented in Section 4.2.

Algorithm 3. $\mathbb{SEG}$=*Fusion*($\{S_1, S_2, \ldots, S_n\}$, $\{\mathbb{SEG}_1, \mathbb{SEG}_2, \ldots, \mathbb{SEG}_n\}$)

(1) initialize global segmentation set $\mathbb{SEG} \leftarrow \emptyset$
(2) initialize set of candidates with the result of all individual segmentation processes $\mathbb{C} = \bigcup_{i=1}^{n} \mathbb{SEG}_i$
(3) WHILE $\mathbb{C} \neq \emptyset$
  (a) choose segment $[I, J]$ from $\mathbb{C}$ with maximum global periodicity score
  (b) remove segment from candidates $\mathbb{C} = \mathbb{C} - \{[I, J]\}$
  (c) add segment to global segmentation $\mathbb{SEG} = \mathbb{SEG} \cup \{[I, J]\}$
  (d) search for partial overlaps between $[I, J]$ and any other segment in $\mathbb{C}$
  (e) create new difference segments by eliminating all partial overlaps
  (f) test all difference segments for periodicity and length
  (g) update $\mathbb{C}$ by including the successfully tested difference segments
(4) END WHILE

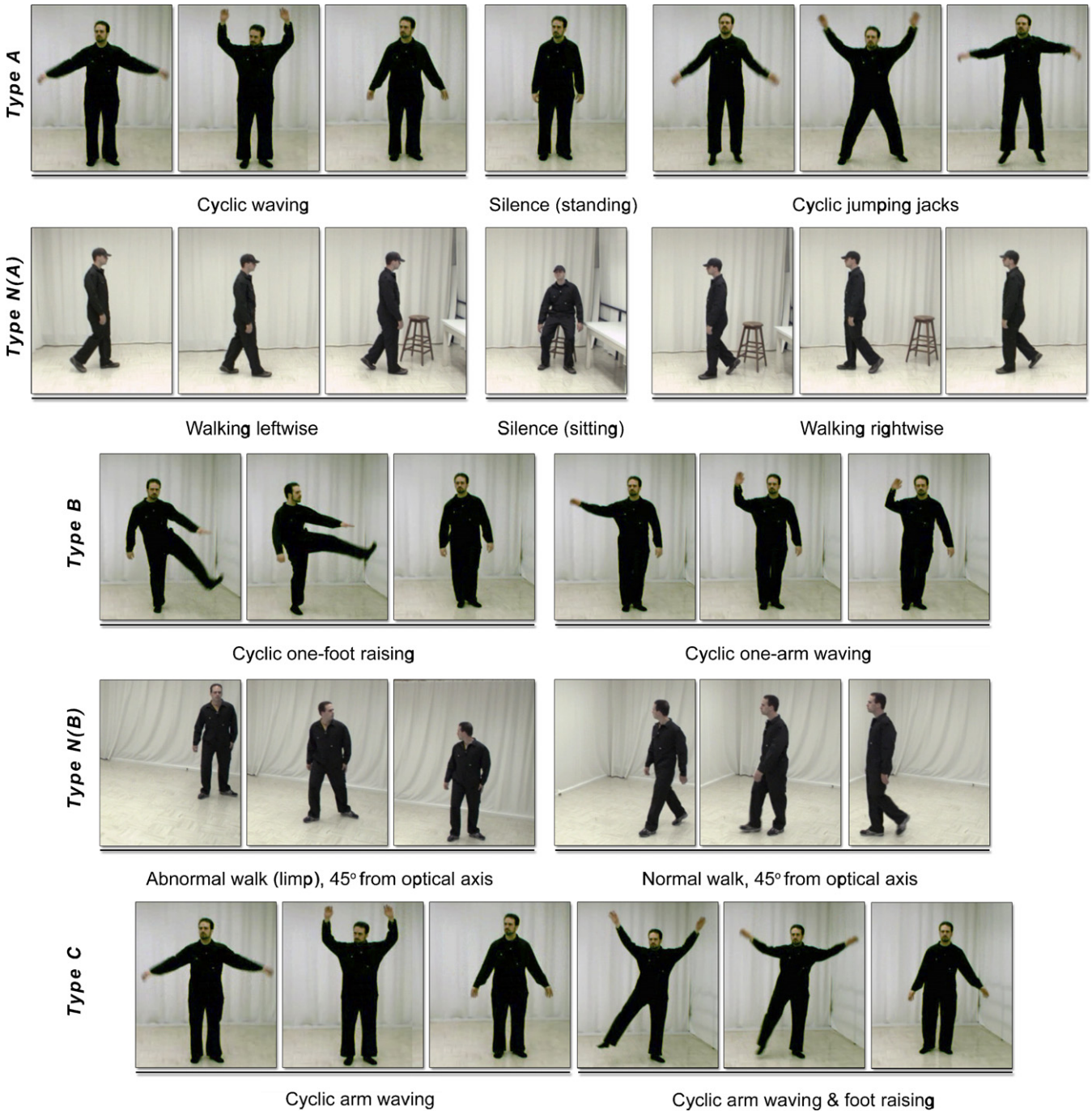

Fig. 6. Examples of sequences of A, B, N, and C types containing ample and natural motion.

4.1. Design of the experiment

The database for this study contains sequences with cyclic activities involving ample limb motion (arm waving, side-stepping, and various combinations of synchronized arm and leg motion), as well as sequences containing natural cyclic motion (walking). The video sequences were acquired with a monocular camera in front of a static background at 30 frames per second; they contain between 2 and 5 cyclic activities each and their total length varies between 300 and 1200 frames. In sequences containing natural cyclic motion, walking along different linear trajectories is interpreted as different cyclic activities; such an interpretation serves well the practical purpose of detecting and analyzing changes of direction in the trajectory of pedestrians. Moreover, the capacity of the proposed approach to differentiate between normal and abnormal gaits was also tested (see Fig. 6).

The test sequences are partitioned according to their expected level of difficulty. Type A sequences contain cyclic activities temporally bounded by pauses or silences. In sequences of type

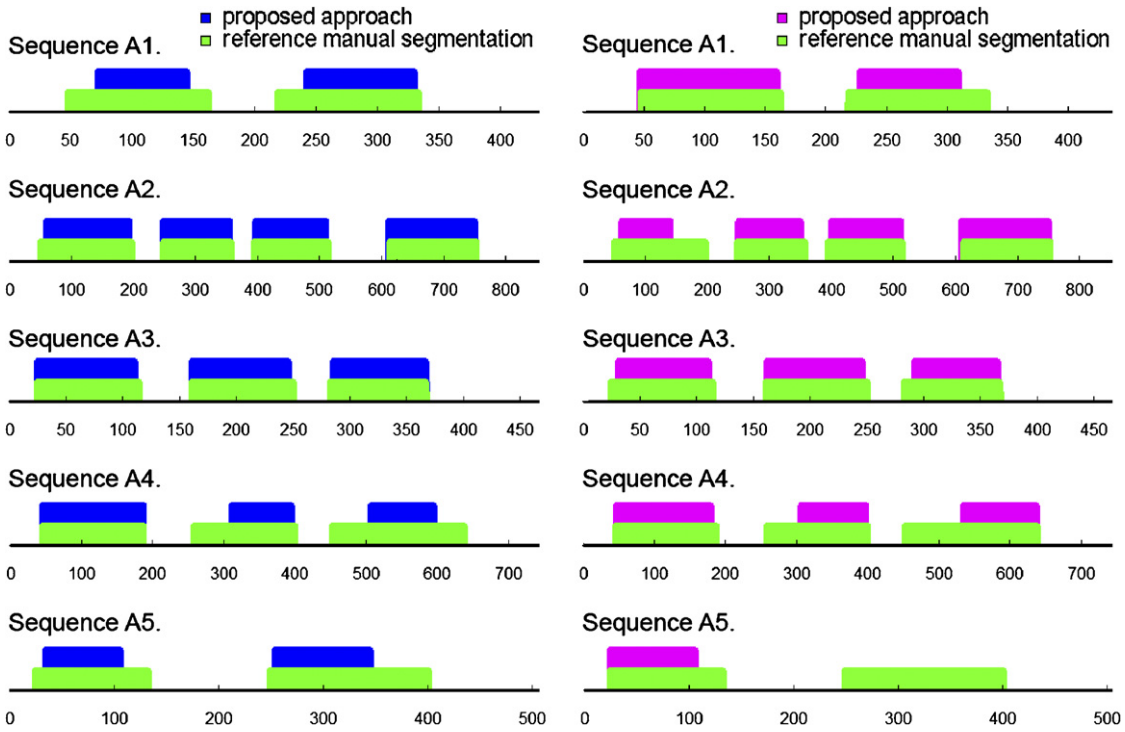

Fig. 7. Segmentation results for A-type sequences. Left: input signals obtained with SPD1; right: input signals obtained with SPD2.

B, at least one activity is temporally adjacent to another activity or to non-cyclic movements. Finally, in sequences of type C at least one activity fuses with another, like waving one arm immediately followed by waving two arms. Fig. 6 presents examples from each sequence type. Test sequences of type A and B which contain natural cyclic motions are referred to as N-type sequences.

The design of the experiment involves two simplifying assumptions. First, a cyclic activity must contain at least three cycles in order to be detected; this constraint is helpful for eliminating false detections due to noisy input signals. Second, it is assumed that the maximum frequency of a cyclic activity is 5 Hz. These assumptions result in $\beta = 18$, where β stands for the minimum length of a cyclic activity.

4.2. Performance analysis and validation

The periodicity thresholds η_l and η_h used in the individual and global signal segmentation steps have been determined empirically using a thorough performance analysis of the proposed approach against manual reference segmentation over the entire database. The selected values are $\eta_l = 45\%$ and $\eta_h = 85\%$. They provided an optimal performance of the proposed approach on our test sequences. Moreover, our approach yields stable results when $\eta_l \in [15\%, 60\%]$ and $\eta_h \in [75\%, 85\%]$.

The performance of the proposed approach was measured using as a reference the average manual segmentation from ten volunteers who outlined the temporal boundaries of cyclic activities. The validation results can be visualized in Figs. 7–9. In addition, two quantitative measures, precision and recall, are used to compare the obtained segmentation with respect to the corresponding reference segmentation on a sequence-by-sequence basis. Precision and recall measures help determine whether the obtained segmentation is sufficiently accurate. A true positive corresponds to a detected segment for which $\tau\%$ of its length overlaps a reference periodic segment. In all experiments τ is set to 75%. False positives correspond to segments with either no such correspondence or with correspondence with an already assigned reference segment. For each analyzed sequence, precision is the ratio of the number of true positives to the total number of detected segments. Recall is the ratio of the number of true positives to the total number of periodic segments in the reference.

Tables 1–3 summarize the results of the validation process. It includes start and end frame numbers of cyclic activities, as detected with our approach and from the corresponding reference segmentation, as well as the computed precision and recall for each sequence containing ample limb motion. Table 4 contains the same information as Tables 1–3 for the test sequences containing natural motion. Due to previously mentioned limitations of SPD1, the detection of significant points for natural motion was performed only with SPD2.

Tables 1–4, together with Figs. 7–9, indicate that the proposed approach performs well. More than three out of four experiments resulted in perfect precision and recall. A majority of the remaining cases have perfect precision which means no false detection. Recall and precision are always at 50% or more. Only three sequences had false detections, two of which also have missed detections. Missed and false detections result

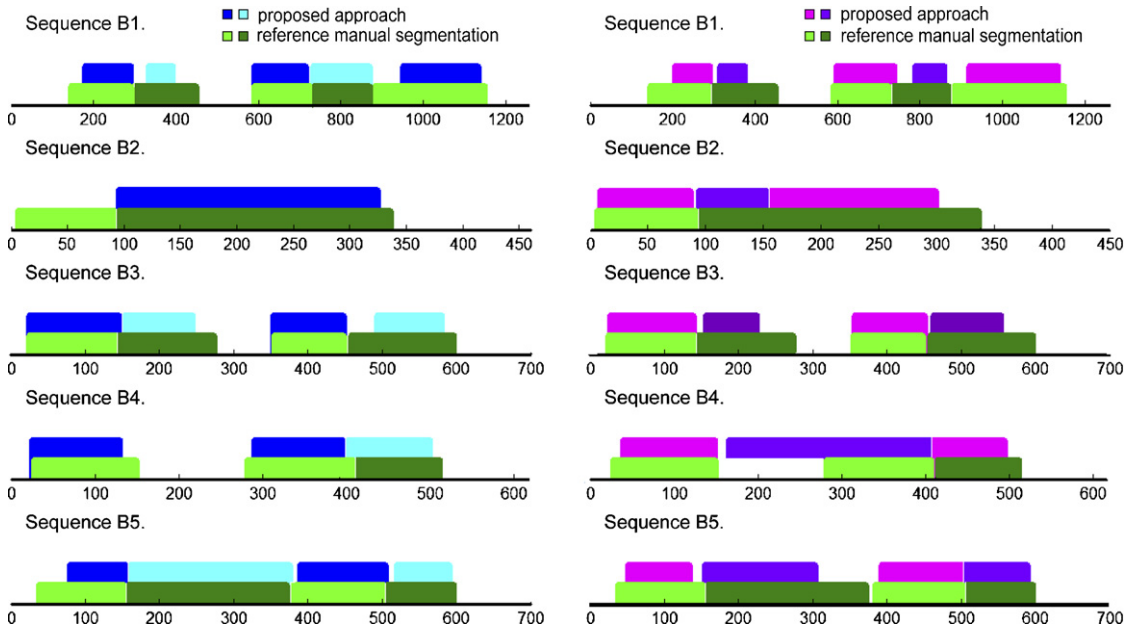

Fig. 8. Segmentation results for B-type sequences. Left: input signals obtained with SPD1; right: input signals obtained with SPD2.

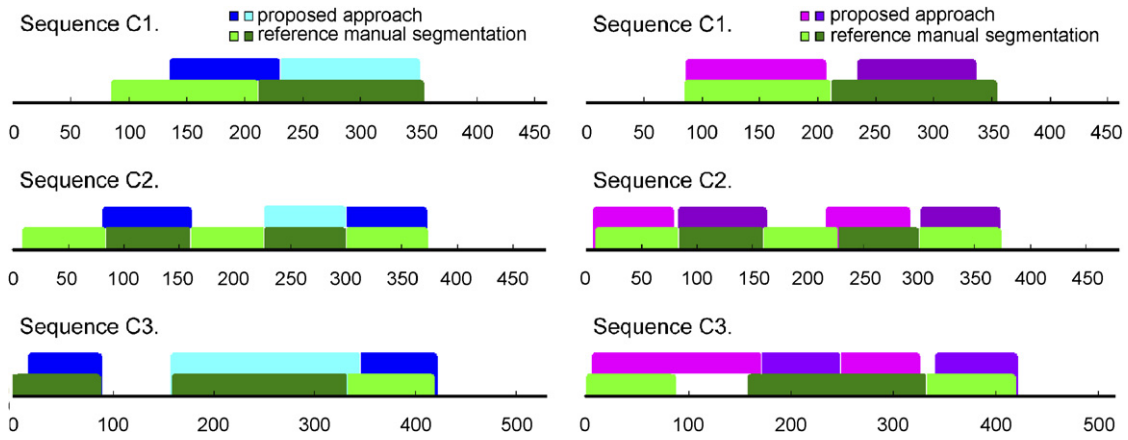

Fig. 9. Segmentation results for C-type sequences. Left: input signals obtained with SPD1; right: input signals obtained with SPD2.

from using fixed periodicity thresholds η_l and η_h over the entire database.

A missed detection from natural cyclic motion sequence N5 is shown in Fig. 10b and c. This sequence contains four cyclic activities defined by four different walking directions with respect to the camera axis, namely: orthogonal to the axis and left-wise; parallel to the axis and away from the camera; orthogonal and right-wise; parallel and towards the camera. The missed detections correspond to the parallel trajectories of motion featuring a low signal-to-noise ratio due to low-amplitude signals. Walking away or towards the camera along the camera axis is not detectable as a cyclic activity using our approach. However, other walking directions (90°, 45° from the optical axis) always gave rise to successful detections in sequences N1–N5.

Differences may exist between the obtained start and end frame numbers and their correspondents in the reference manual segmentation even when a cyclic activity is properly detected. Those differences occur since factors in the way humans perceive periodicity are not accounted for by the proposed periodicity measure. Indeed, human perception may accommodate large variations in speed, amplitude and frequency between successive cycles of the same activity; the proposed approach tolerates only a limited amount of inter-cycle variability, with the upper limit fixed by the threshold η_l. Besides, one may recall that the reference segmentation is an average which may partly explain the noted differences.

Finally, one may ask which set of signals (generated for significant points detected with SPD1 or SPD2) is more suitable

Table 1
Experimental results obtained for A-type sequences

Sequence (no. of contained activities)	Start-end frames (proposed approach)		Start-end frames (reference)	Recall (%)		Precision (%)	
	SPD1	SPD2		SPD1	SPD2	SPD1	SPD2
A1 (2)	(71–146; 242–331)	(46–161; 228–310)	(47–163; 218–334)	100	100	100	100
A2 (4)	(58–195; 246–357; 396–512; 610–752)	(61–142; 249–353; 398–514; 608–752)	(48–199; 247–359; 393–516; 612–755)	100	100	100	100
A3 (3)	(24–111; 160–247; 285–368)	(30–111; 161–247; 291–366)	(24–115; 160–251; 282–368)	100	100	100	100
A4 (3)	(45–189; 310–397; 506–597)	(45–181; 305–399; 533–639)	(44–189; 257–402; 451–640)	100	100	100	100
A5 (2)	(34–107; 253–346)	(24–107)	(24–134; 248–402)	100	100	50	100

Input data are extracted with SPD1 and SPD2.

Table 2
Experimental results obtained for B-type sequences

Sequence (no. of contained activities)	Start-end frames (proposed approach)		Start-end frames (reference)	Recall (%)		Precision (%)	
	SPD1	SPD2		SPD1	SPD2	SPD1	SPD2
B1 (5)	(178–294; 332–395; 588–718; 732–873; 950–1136)	(204–293; 314–378; 595–741; 786–862; 918–1138)	(145–292; 297–452; 588–730; 733–873; 881–1152)	100	100	100	100
B2 (2)	(95–325)	(8–88; 94–154; 155–300)	(6–93; 95–337)	50	100	100	67
B3 (4)	(24–147; 152–246; 353–450; 493–582)	(26–141; 155–226; 356–452; 462–555)	(24–143; 145–275; 354–450; 458–598)	100	100	100	100
B4 (3)	(24–131; 290–398; 402–501)	(39–150; 165–407; 408–496)	(27–151; 281–409; 413–513)	100	100	67	67
B5 (4)	(78–159; 160–377; 389–506; 519–592)	(50–135; 154–304; 392–503; 504–591)	(37–154; 158–373; 383–504; 506–598)	100	100	100	100

Input data are extracted with SPD1 and SPD2.

Table 3
Experimental results obtained for C-type sequences

Sequence (no. of contained activities)	Start-end frames (proposed approach)		Start-end frames (reference)	Recall (%)		Precision (%)	
	SPD1	SPD2		SPD1	SPD2	SPD1	SPD2
C1 (2)	(137–229; 233–349)	(88–205; 236–335)	(87–210; 213–353)	100	100	100	100
C2 (5)	(82–159; 229–299; 300–371)	(9–77; 85–161; 218–290; 304–371)	(11–82; 83–159; 160–226; 228–298; 302–372)	60	100	80	100
C3 (3)	(18–86; 159–344; 345–420)	(8–170; 171–247; 248–324; 343–419)	(2–85; 160–331; 333–417)	100	100	67	50

Input data are extracted with SPD1 and SPD2.

for detecting and differentiating between cyclic activities. For a better visualization of this comparison, Figs. 7–9 may be displayed side by side. Some sequences are better segmented when using input data extracted with SPD1 (A5, B4, C3), while SPD2 works better for other sequences (C2 and perhaps B2); in general, the performances of the two methods are comparable. The similar quality of the segmentations obtained with input data extracted with either SPD1 or SPD2 is an encouraging result; it indicates that natural cyclic motion can be successfully detected using a small set of signals extracted with real-time tracking.

4.3. Computational complexity

The individual segmentation step computes the length-normalized score for every possible segment $[i, j]$ at each iteration. For a signal S of length l_S, there are $(l_S - \beta)^2/2$ possible segments with a minimum length of β. The computational complexity for an individual periodicity score is $\mathcal{O}(l_S \log(l_S))$ if the autocorrelation is computed using the fast Fourier transform (FFT). Therefore, the computational complexity of the individual segmentation is $\mathcal{O}(l_S \log(l_S) \cdot (l_S - \beta)^2)$ which reduces to $\mathcal{O}(l_S^3 \log(l_S))$ when β is small with respect to l_S, as is usually the case. The length l_S of all signals in the signal set representing a video sequence is equal to the length of the sequence.

The global segmentation step has two computationally intensive components. The first one consists in the pre-computation of global periodicity scores for all candidates extracted during individual segmentation. The computation of one global periodicity score has complexity $\mathcal{O}(nl_S \log(l_S))$ where n is the number of signals in the set ($n = 34$ for SPD1 and $n = 22$ for SPD2). The maximum number of candidates for a given signal set is less than nl_S/β, since one signal cannot contain more than l_S/β segments. Therefore, the computation of periodicity scores for all candidates is bounded by $\mathcal{O}(n^2 l_S^2 \log(l_S))$.

The second computationally intensive component of the global segmentation is the iterative test for updating the set of candidates with newly created difference segments (see pseudo code of Algorithm 3). This test is performed during a maximum number of $n^2(l_S/\beta)^2$ iterations; therefore, its computational complexity is limited by $\mathcal{O}(n^2 l_S^3 \log(l_S))$. One may conclude that the global segmentation step has a computational complexity of $\mathcal{O}(n^2 l_S^3 \log(l_S))$.

The proposed approach for temporal segmentation was implemented on a 3.0 GHz Pentium IV personal computer with 1024 MB RAM. The time necessary for performing the temporal segmentation on test sequences in the database varies between 2.5 s and 5 min, depending on the length of the sequence and its content. Approximations to the individual segmentation step, which strongly dominates the computation time, is therefore needed in order to limit the computation time of the algorithm for near real-time applications. Such approximations are currently under study for a comparative performance evaluation.

Table 4
Experimental results obtained for N-type sequences

Sequence (no. of contained activities)	Start-end frames (proposed approach)	Start-end frames (reference)	Recall (%)	Precision (%)
N1 (2)	(0–154; 378–518)	(0–154; 369–519)	100	100
N2 (2)	(6–176; 249–416)	(0–195; 235–460)	100	100
N3 (2)	(11–133; 215–343)	(0–140; 170–360)	100	100
N4 (2)	(38–121; 128–259)	(25–140; 141–260)	100	100
N5 (4)	(22–165; 415–560)	(0–153; 173–373; 391–555; 570–810)	50	100

Input data are extracted with SPD2.

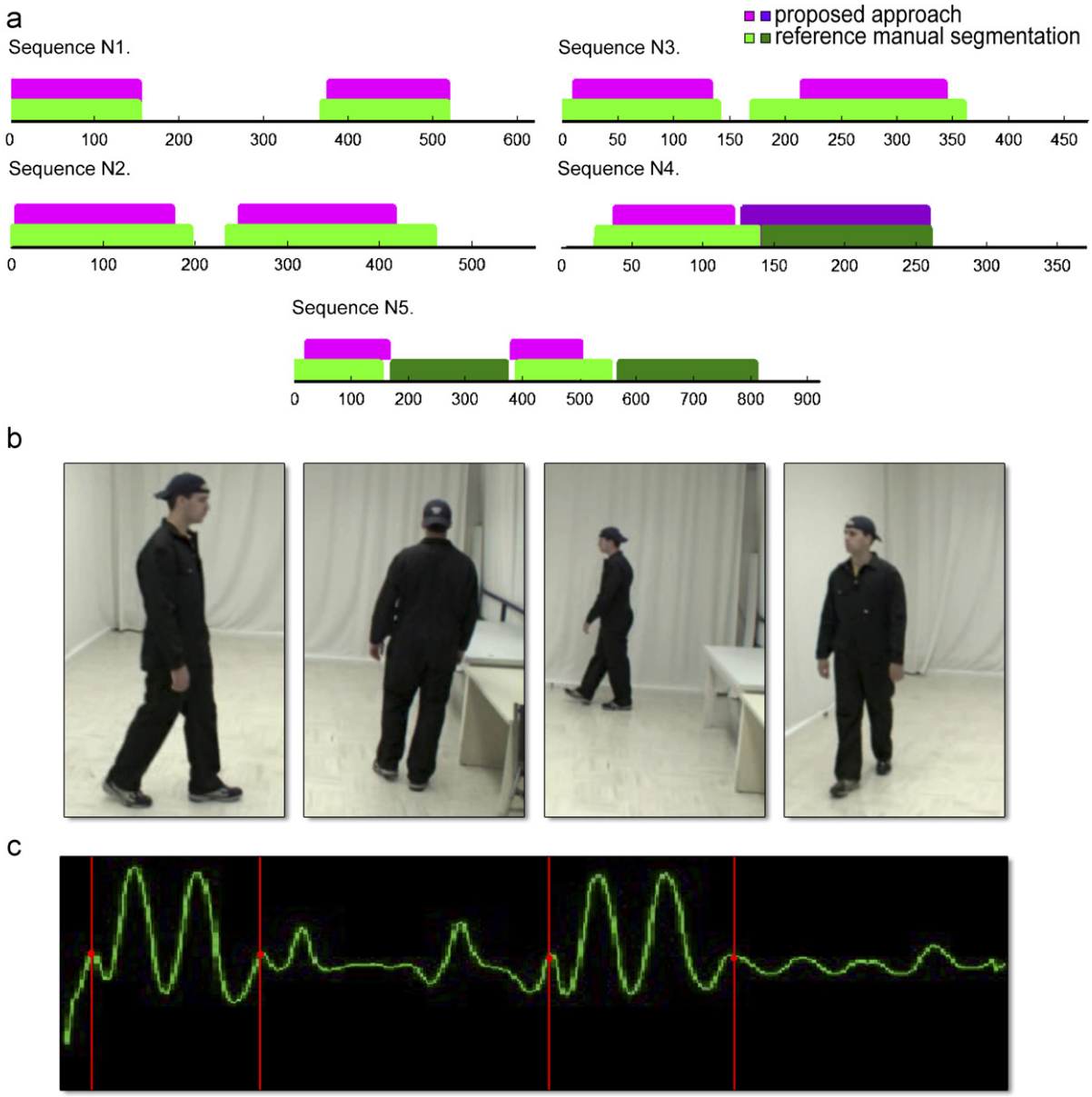

Fig. 10. (a) Segmentation results for N-type sequences; (b) key frames in sequence N5 representing different cyclic actions; and (c) identification of periodic portions on one signal in the input set for sequence N5.

5. Conclusions

This paper has presented a new approach for the temporal segmentation of cyclic activities using multiple trajectories of body parts. These trajectories were extracted using two different methods and assembled into a set of 1-D signals which represents the input data for the proposed segmentation approach. The rationale behind the chosen data representation is the direct correspondence between a cyclic human activity and periodic segments located on 1-D signals. Periodicity information is first extracted on a signal-by-signal basis using a length-normalized periodicity score and a greedy algorithm. This first step identifies on each signal which segments are most likely to indicate cyclic activities. A second step combines individual detections into a global segmentation using a global periodicity score and a maximum search algorithm which updates the pool of candidates iteratively.

The proposed approach has been successfully tested on a variety of sequences containing cyclic activities such as aerobic exercises and walking along different directions. The validation has also proved the robustness of the proposed approach with respect to the way the input data (i.e. the set of signals describing the sequence of activities) is generated. Experimental results indicate that natural cyclic motion can be successfully detected using a small set of signals describing head, hands and feet motion and extracted with real-time tracking.

This paper advances the state-of-the-art in video-based human motion analysis by filling a missing link in the video understanding process. This missing link corresponds to the accurate detection of temporal limits of the activities of interest within a video stream. As outlined in the introduction, it is believed that the temporal segmentation of an activity is an essential step for activity representation and recognition. It was shown that this temporal segmentation is feasible for human cyclic activities of different levels of complexity.

Ongoing work focuses on the reduction of the rates of false and missed detections by optimizing the global segmentation step; a greedy approach might not be ideal as false positives in the individual signal segmentation are likely to survive at the next step. Also, future work will explore various other approaches for extracting the set of input signals, in order to improve the signal-to-noise ratio in the input data.

Acknowledgments

This work is supported by FQRNT through a postgraduate scholarship and by NSERC discovery grants.

References

[1] R. Polana, R. Nelson, Low level recognition of human motion, in: Proceedings of IEEE Workshop on Motion of Non-rigid and Articulated Objects, Austin, TX, USA, 1994, pp. 77–82.

[2] J. Ben-Arie, Z. Wang, P. Pandit, S. Rajaram, Human activity recognition using multidimensional indexing, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 1091–1104.

[3] Y. Rui, P. Anandan, Segmenting visual actions based on spatio-temporal motion patterns, in: Proceedings of IEEE International Conference on Computer Vision Pattern Recognition (CVPR2000), Hilton Head Island, SC, USA, 2000, pp. 111–118.

[4] J. Gao, A.G. Hauptmann, H.D. Wactlar, Combining motion segmentation with tracking for activity analysis, in: Proceedings of International Conference on Automatic Face and Gesture Recognition (FGR04), Seoul, Korea, 2004, pp. 699–704.

[5] J. Min, R. Kasturi, Extraction and temporal segmentation of multiple motion trajectories in human motion, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR2004), Washington, DC, USA, 2004, pp. 118–122.

[6] S. Quirion, A.B. Albu, R. Bergevin, Skeleton-based temporal segmentation of human activities from video sequences, in: Proceedings of the 13th International Conference in Central Europe on Computer Graphics (WSCG 05), Plzen-Bory, Czech Republic, 2005, pp. 145–148.

[7] A. Plotnik, S. Rock, Quantification of cyclic motion of marine animals from computer vision, in: Proceedings of the MTS/IEEE Oceans 2002, vol. 3, Biloxi, MS, USA, 2002, pp. 1575–1581.

[8] R. Polana, R. Nelson, Detection and recognition of periodic, non-rigid motion, Int. J. Comput. Vis. 23 (1997) 261–282.

[9] R. Cutler, L. Davis, Robust real-time periodic motion detection, analysis, and applications, IEEE Trans. Pattern Anal. Mach. Intell. (2000) 781–796.

[10] Y. Liu, R. Collins, Y. Tsin, Gait sequence analysis using frieze patterns, in: Proceedings of the Seventh European Conference on Computer Vision (ECCV'02), Copenhagen, Denmark, 2002, pp. 657–671.

[11] A. Thangali, S. Sclarroff, Periodic motion detection and estimation via space-time sampling, in: IEEE Workshop on Motion and Video Computing, Breckenridge, CO, USA, 2005, pp. 176–182.

[12] H. Lin, L. Wang, S.N. Yang, Extracting periodicity of a regular texture based on autocorrelation functions, Pattern Recognition Lett. 18 (1997) 333–343.

[13] Y. Ran, Q. Zheng, I. Weiss, L.S. Davis, W. Abd-Almageed, L. Zhao, Pedestrian classification from moving platforms using cyclic motion pattern, in: International Conference on Image Processing, (ICIP05), Genova, Italy, 2005, pp. 854–857.

[14] Y. Ran, I. Weiss, Q. Zheng, L. Davis, An efficient and robust human classification algorithm using finite frequencies probing, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR2004), Washington, DC, USA, 2004, pp. 132–136.

[15] J. Little, J. Boyd, Recognizing people by their gait: the shape of motion, Videre: J. Comput. Vis. Res. 1 (1998) 24–42.

[16] P. Tsai, M. Shah, K. Keiter, K. Kasparis, Cyclic motion detection, Pattern Recognition 27 (1994) 1591–1603.

[17] D. Cunado, M. Nixon, J. Carter, Automatic extraction and description of human gait models for recognition purposes, Comput. Vis. Image Understanding 90 (2003) 1–41.

[18] M. Yazdi, A. Branzan-Albu, R. Bergevin, Morphological analysis of spatiotemporal patterns for the temporal segmentation of cyclic activities, in: Proceedings of International Conference on Pattern Recognition (ICPR04), Cambridge, UK, 2004, pp. 240–243.

[19] J. Vignola, J.-F. Lalonde, R. Bergevin, Progressive human skeleton fitting, in: Proceedings of the 16th Vision Interface Conference, Halifax, Canada, 2003, pp. 35–42.

[20] F. Jean, R. Bergevin, A. Branzan-Albu, Body tracking in human walk from monocular video sequences, in: Second IEEE Canadian Conference on Computer and Robot Vision (CRV 2005), Victoria, Canada, 2005, pp. 144–151.

[21] C.D.S. Seitz, View invariant analysis of cyclic motion, Int. J. Comput. Vis. 25 (1997) 231–251.

About the Author—A. BRANZAN ALBU received the Ph.D. degree from the Polytechnic Institute of Bucharest in 2000. In 2001, she joined the Computer Vision and Systems Laboratory at Laval University as a Postdoctoral Researcher and became an Assistant Professor at Laval in 2003. In 2005, she joined the ECE Department at the University of Victoria (BC). Her research interests include computer vision-based human motion analysis and medical imaging. Dr. Branzan Albu is a member of the Province of British Columbia Association of Professional Engineers (APEGBC).

About the Author—R. BERGEVIN received the Ph.D. degree in Electrical Engineering from McGill University in 1990. He joined the Computer Vision and Systems Laboratory at Laval University in 1991. His research interests are in image analysis and cognitive vision. Dr. Bergevin is a Member of the Province of Quebec's Association of Professional Engineers (OIQ) and the IEEE Computer Society. He serves as Associate Editor for the Pattern Recognition journal and Area Editor for the Computer Vision and Image Understanding journal.

About the Author—S. QUIRION received the M.Sc. degree from Laval University specializing in Computer Vision in 2006. He has also received the B.Sc. degree from Laval University in Computer Science in 2003. He is currently pursuing Doctoral studies specializing in automated motor learning and realistic motion synthesis.