Morphological Analysis of Spatio-Temporal Patterns for the Segmentation of Cyclic Human Activities

Mehran Yazdi, Alexandra Branzan Albu, and Robert Bergevin
Computer Vision and Systems Laboratory
Dept of ECE, Laval University, Québec, Canada, G1K 7P4
{*yazdi*, *branzan*, *bergevin*}@gel.ulaval.ca

Abstract

This paper describes a new method for the temporal segmentation of human actions based on a 2D inter-frame similarity plot. This similarity matrix contains relevant information for the analysis of cyclic and symmetric human activities, where the motion performed during the first semi-cycle is repeated in the opposite direction during the second semi-cycle. Thus, the pattern associated to a periodic activity in the similarity matrix is rectangular and decomposable into elementary units. We propose a morphology-based approach for the detection and analysis of activity patterns. Pattern extraction is further used for the detection of the temporal boundaries of the cyclic symmetric activities. Result evaluation approach is based on a statistical estimation of the ground truth segmentation and on a confidence ratio for temporal segmentations. Research reported in this paper was supported by a discovery grant of the National Sciences and Engineering Research Council of Canada.

1. Introduction

In the context of human motion analysis, the cyclic aspect is mainly used as a cue for detecting activities such as walking, running, and target identification through gait recognition. Tsai *et al* [1] detect walking using the spatio-temporal curvature function of trajectories corresponding to specific points on the human target in motion. Their technique is designed for motion-based recognition, namely for identifying the tracked object from its motion.

Polana and Nelson [2][3] introduce the concept of temporal texture for the detection of periodic activities such as walking, exercise, and swing. Seitz and Dyer [4] define the notion of period trace which allows to relax the assumption that a motion should be perfectly even from one cycle to the next.

While the study of sequences dedicated to a single activity of interest has led to interesting results in human motion representation, there is little research about video sequences where the activity pattern changes over time. Recent work by Bobick and Davis [5] deals with the temporal segmentation of video sequences into actions based on a backward-looking temporal time window.

In this paper, we describe a new method for the temporal segmentation of human activities based on the 2D inter-frame similarity plot introduced by Cutler and Davis in [6]. In the context of our research, we have investigated the relevance of the 2D inter-frame similarity plot for the detection of cyclic and symmetric actions from video sequences containing multiple activities.

The rest of the paper is organized as follows. Section 2 contains a description of our approach. The experimental results are discussed in section 3. Section 4 contains the conclusions as well as the future work directions.

2. An overview of the proposed approach

The modular diagram of the proposed temporal segmentation approach is shown in Fig. 1. The pre-processing phase contains three steps: the background subtraction, the shadow removal, and the silhouette normalization. After preprocessing, the initial sequence is transformed into a sequence of binary images containing human silhouettes in bounding boxes of standard size. The inter-frame similarity matrix computed by cross-correlation is displayed as an image. This image represents input data for our morphology-based approach designed for the detection and analysis of spatio-temporal patterns corresponding to periodic human activities.

2.1. Preprocessing

The *background subtraction* consists of classifying the pixels in each frame of the sequence as belonging either to

background or to the foreground. Since our database is acquired with a static camera in an indoor environment, we deal with a static background and compute the pixel-by-pixel difference between every frame and the reference image of the background. Then, the difference image is binarized using the Otsu thresholding. Median filtering eliminates artifacts due to noise and specular effects.

Since the background subtraction does not eliminate shadows in an environment with uncontrolled lighting, further processing for *shadow removal* is needed.

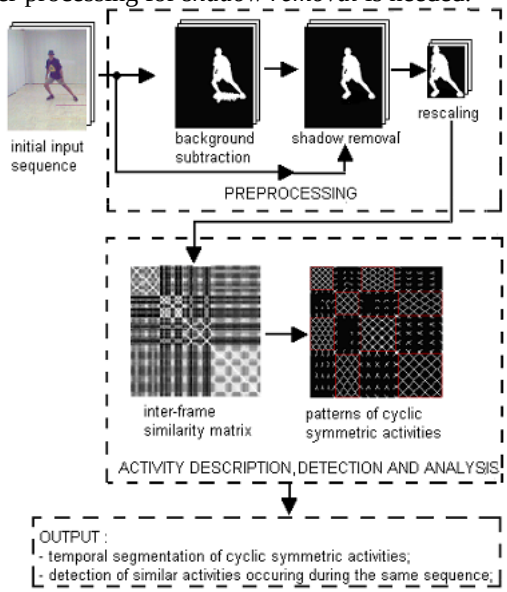

Figure 1. Modular diagram of the proposed approach

The shadow of a moving person exhibits similar chromaticity but lower brightness than the corresponding background region. Thus, the foreground silhouette is first partitioned into a set of subregions, using automatic mode separation in the intensity histogram. Mode separation is performed by finding the local minimums with standard signal analysis techniques. Every subregion of the silhouette contains only pixels located in the same histogram mode. Next, we determine which subregions in the silhouette correspond to shadow. In the HSV (Hue-Saturation-Value) colour space, the chromatic appearance of the region is represented by the $[\mu_H, \sigma_H, \mu_S, \sigma_S]$ vector, where μ_H and μ_S are the average values of the hue and saturation, while σ_H and σ_S are the standard deviations respectively. Each subregion in the silhouette is compared with its corresponding subregion in the background reference image by means of Euclidian distance between the feature vectors. Our method uses a threshold related to the acceptable level of similarity between a subregion of the silhouette and the corresponding background. The threshold value is not adjustable from one frame to another and is constant for a given sequence. Any subregion exceeding this similarity threshold is considered to be shadow and it is removed.

Silhouette normalization is necessary for obtaining similar descriptions of the same activity performed at different depths relative to the camera and to compensate for depth changes occurring between successive cycles of the same activity. Bounding boxes corresponding to the binary blobs are rescaled to a standard size using the nearest neighbor interpolation method.

2.2. Activity description, detection, and analysis

Cutler and Davis [6] have introduced the notion of 2D inter-frame similarity plot for the time-frequency analysis of periodic motion. We propose a new analysis perspective and redefine the previous concept as follows. Given a sequence of N binary normalized frames, the inter-frame similarity matrix is $\left\lfloor r_{ij} \right\rfloor_{1 \le i \le N, 1 \le j \le N}$, where r_{ij} is the correlation of frames i and j. Binary inter-frame cross-correlation is more robust than its gray-level equivalent with respect to specular effects due to ambient lighting. The display of the similarity matrix maps the [-1,1] range of the correlation coefficients onto a 256 gray-level image, as it is shown in Fig. 2b.

The similarity matrix exhibits some relevant properties for the analysis of a particular category of human activities. Specifically, we are interested in detecting periodic and symmetric activities, where the motion performed during the first semi-cycle is repeated in the opposite direction during the second semi-cycle. In a cyclic symmetric activity, consecutive frames belonging to the first semi-cycle and similar to the reference frame (i.e. the first frame in the row) form bright segments parallel to the main diagonal (see Fig. 2a, center). In addition, bright segments orthogonal to the main diagonal represent the second semi-cycle. A periodic and symmetric activity is represented by a zigzag pattern where the primitive is a V shape corresponding to one cycle (see Fig. 2a, left). *As shown in Fig. 2a, the pattern associated to a periodic activity in the inter-frame similarity matrix is rectangular, and can be further decomposed into elementary units.*

The first goal is to isolate the activity patterns. First, the bright regions are extracted by thresholding at 60% of the maximum brightness (see Fig. 2c). Next, a sequence of standard morphological operators is used for pattern extraction: a) 2 iterative dilations with a 5 pixel-sized cross-shaped structuring element removes possible line disconnections; b) shrinking reshapes the dilated image edges into one-pixel thick sets of linear segments, while preserving the connectivity of the lines; c) isolated pixels and short line segments are removed. The result of the

morphological processing (see Fig. 2d) contains separable patterns corresponding to cyclic activities respectively.

Pattern extraction is performed with a standard region growing technique, based on the fact that the pattern corresponding to a cyclic activity can be seen as a set of adjacent elementary closed contours (see Fig. 2a, right). Thus, elementary regions are grown inside the closed contours (see Fig. 2e, where a colour display is used for differentiating adjacent elementary regions). Next, neighboring regions are merged to form activity patterns.

After pattern extraction, motion information captured within the patterns is analyzed. First, every cyclic symmetric activity in the input sequence has a corresponding pattern aligned on the main diagonal of the similarity matrix. Therefore, the upper-left and lower-right corners of the bounding box enclosing the targeted pattern correspond to the temporal boundaries of the activity (see Fig. 2f). Results on temporal segmentation will be presented and evaluated in the next section. Second, the number of the cycles in the activity is related to the number of elementary regions in the pattern. Third, patterns not aligned on the main diagonal correspond to similar activities performed at different moments during the same sequence (see Fig. 2g).

3. Experimental results

The database of this study consists in eight video sequences acquired with a monocular camera in an indoor office environment at a frame rate of 30 frames/second. The frame size is 480x640 pixels, while the length of the sequences varies between 170 and 670 frames. Each video sequence in the database corresponds to scenarios in which one human subject performs several activities, such as cyclic aerobic exercises (arm swinging, arm waiving, leg bending, and combinations of arm and leg motions) in alternation with walking, standing, sitting etc. Fig. 3 contains a collection of relevant frame samples belonging to a video sequence where four cyclic and symmetric actions occur.

We aim at an accurate temporal extraction of the cyclic and symmetric human activities in each input sequence. Our approach has successfully detected every cyclic symmetric activity in the database, as well as its temporal boundaries. In order to evaluate our temporal segmentation approach, we need to estimate ground truth segmentation, since there is a non-negligible inter-observer variability in human segmentation. A statistical estimation of the ground truth was built from ten human segmentations, which have independently marked the activity boundaries in the database sequences. The distribution of human segmentations for a given sequence is outlined as a histogram of action boundaries versus the frame index. The histogram of action boundaries detected by the human observers for the sequence in Fig. 3 is shown in Fig. 4. The left and right histogram maxima provide an accurate ground truth estimate for the boundaries of every detected activity. Next, we evaluate the performances of our approach with respect to the estimated ground truth (EGT) segmentation.

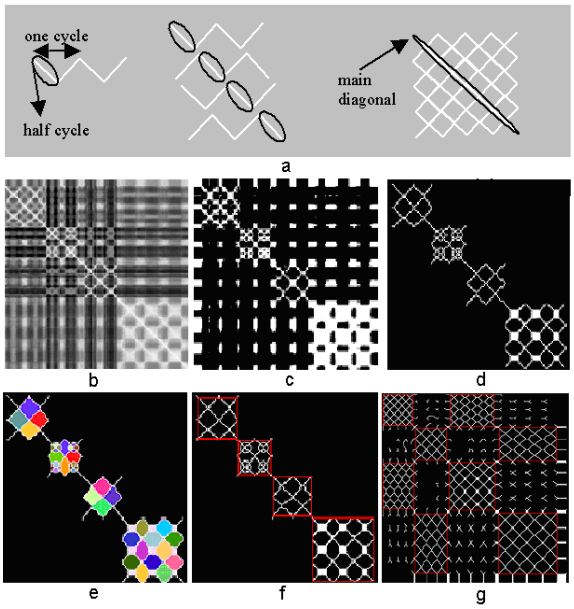

Figure 2. a) pattern forming in the similarity matrix; b) similarity matrix; c) thresholding; d) morphological processing; e) pattern extraction with region growing; f) bounding boxes for patterns centered on the main diagonal; g) bounding boxes for every pattern in the image

For a quantitative evaluation for a given temporal segmentation *S*, we define a *confidence ratio* as follows:

$$C(i,S)=\begin{cases} N(i)/10 & \text{if } S(i)=true \\ 1-N(i)/10 & \text{otherwise} \end{cases}$$

where $S(i)$ is true if and only if frame i is detected by the evaluated segmentation S as part of the activity; $N(i)$ is the number of human observers that have also detected frame i as belonging to the activity. The normalization coefficient is set to 10 since 10 human manual segmentations where used in the estimation. The confidence ratio C takes values in the interval [0, 1]; low values of $C(i, S)$ mean that few observers have taken the same decision for frame i as S. On the contrary, high values of $C(i, S)$ mean that the majority of observers agree with S. The previously defined evaluation measure allows for comparing the automatic segmentation performed by the proposed algorithm with the estimated ground truth segmentation. Fig. 4 (right column) shows the plot of the confidence ratio corresponding to the actions in Fig. 3.
Since the confidence ratio C performs a frame-by-frame evaluation, it conveys information about local errors, and about the global segmentation quality as well. Thus, for

both considered segmentations (automatic and EGT), there are transitory phases at the beginning and the end of each action. Since the transitory phases are highly similar for both segmentations in the case of every considered action (see Fig. 4, right column), we conclude that our algorithm yields excellent performances in the segmentation of the considered human activities. Due to the space limit, only cyclic symmetric activities belonging to one sequence were illustrated in Fig. 3 and Fig. 4.

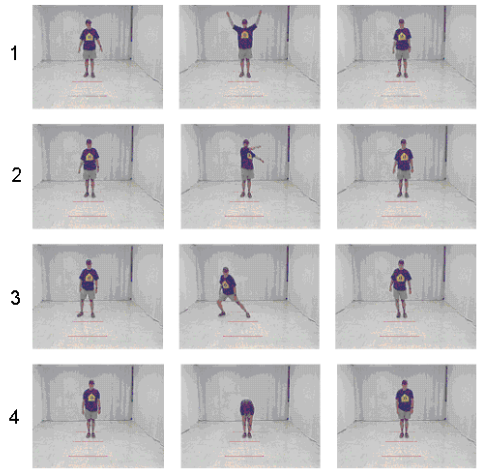

Figure 3. Relevant frame samples in a sequence containing four cyclic and symmetric activities 1) arm waiving; 2) arm rotation; 3) leg flexing; 4) torso bending.

Performance evaluation over the entire database leads to the following conclusions: a) the global average boundary detection error is 4.76 frames, which is rather encouraging at a 30 frames/sec rate; b) maximal errors (14 frames) were obtained for activities finishing with an incomplete cycle. While the human visual system is able to accurately detect an activity where the last cycle is incomplete, our algorithm does not consider the incomplete cycle as part of the activity.

5. Conclusions and future work

This paper deals with segmenting cyclic symmetric human actions from a continuous video sequence. Specifically, we perform the accurate detection of temporal boundaries for the activities of interest. We redefine the concept of inter-frame similarity matrix [6] and propose a new morphology-based method for extracting relevant motion information from this spatio-temporal template. We have tested our approach on a host of periodic and cyclic human activities, and provided robust statistical ground truth estimation for the validation of our results. The quantitative evaluation of the proposed approach is based on a new measure, called the confidence ratio, which allows for a precise performance assessment. Moreover, the confidence ratio will be appropriate for future comparisons of our approach with other temporal segmentation methods. Ongoing work in our project deals with relaxing the symmetry constraint, which will allow the analysis of natural cyclic human actions, such as walking or running.

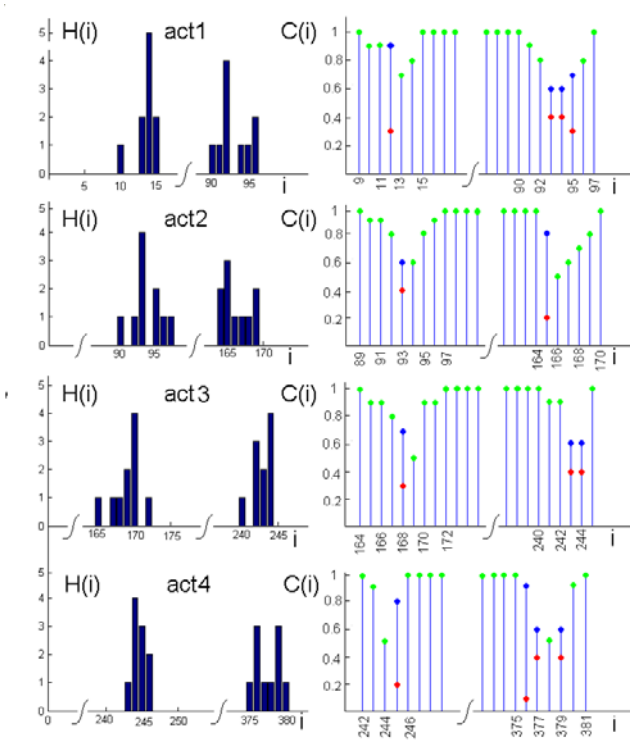

Figure 4. Left: Histograms of the human boundary detections (i is the frame index). Right: confidence ratio versus i for automated and EGT segmentations (• = confounded auto and EGT; • = EGT; • = auto)

6. References

[1] P.-S. Tsai, M. Shah, K. Keiter, and T. Kasparis, "Cyclic motion detection", *Technical Report CS-TR-93-08*, Computer Science Dept, University of Central Florida, 1993.

[2] R. Polana and R. C. Nelson, "Detecting activities", *J. Visual Comm. Image Represenation*, 5(2), 1994, pp. 172-180.

[3] R. Polana and R. Nelson, "Low level recognition of human motion", *IEEE Computer Science Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, TX, 1994, pp. 77-82.

[4] S. M. Seitz and C. R. Dyer, "Detecting irregularities in cyclic motion", In *Proc. Workshop on Motion of Non-Rigid and Articulated Objects*, 1994, pp. 178-185.

[5] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates", *IEEE Trans. on Patt. Anal. and Machine Intell.*, vol. 23, no. 3, 2001, pp. 257-267.

[6] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications", *IEEE Trans. on Patt. Anal. and Machine Intell.*, vol. 22, no. 8, 2000, pp.781-796.