M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 435
Proceedings of the International Conference on Computer Vision,
Bombay, India, January 4-7, 1998

Finding Periodicity in Space and Time *

Fang Liu and Rosalind W. Picard
The Media Laboratory, E15-383
Massachusetts Institute of Technology, Cambridge, MA 02139
fliu@media.mit.edu, picard@media.mit.edu

Abstract

An algorithm for simultaneous detection, segmentation, and characterization of spatiotemporal periodicity is presented. The use of periodicity templates is proposed to localize and characterize temporal activities. The templates not only indicate the presence and location of a periodic event, but also give an accurate quantitative periodicity measure. Hence, they can be used as a new means of periodicity representation. The proposed algorithm can also be considered as a "periodicity filter," a low-level model of periodicity perception. The algorithm is computationally simple, and shown to be more robust than optical flow based techniques in the presence of noise. A variety of real-world examples are used to demonstrate the performance of the algorithm.

1 Introduction

Periodicity is common in the natural world. It is also a salient cue in human perception. Information regarding the nature of a periodic phenomenon, such as its location, strength, and frequency, is important for our understanding of the environment. Techniques for periodicity detection and characterization can assist in many applications requiring object and activity recognition and representation.

Although surface patterns may come to mind first, periodicity often involves both space and time, such as cyclic motion. The main body of work on periodic motion is model-based (e.g., [1][2]). More recently there is work on motion recognition directly using low-level features of motion information (e.g., [3][4][5]). However, to date, there has not been a method which uses low-level features to detect and systematically characterize periodicity in space and time. In this work, we attempt to tackle this problem by using *periodicity templates* to incorporate the location, strength, and other characteristic information of a periodic phenomenon. The templates are useful in applications such as periodic motion representation and action recognition. The template generating procedure provides a tool for detecting and segmenting regions of periodicity. The proposed method is spectral based, and is computationally efficient.

*This work was supported in part by IBM and NEC.

1.1 Our Approach

The approach presented here is motivated by theory for textured image modeling that assumes an underlying random field representation of the data [6]. In particular, 1-D signals along the temporal dimension are considered as stochastic processes. When assuming stationarity, a stochastic signal can be decomposed into deterministic (periodic) and indeterministic (random) components. This is known as Wold decomposition [7]. In the frequency domain, the deterministic and the indeterministic components correspond respectively to the singular and the continuous part of the signal's Fourier spectrum. In practical applications, this is to say that the repetitive structure in the signal contributes only to the spectral harmonic peaks and the random behavior to the smooth part of the spectrum. Therefore, the energy contained in the spectral harmonic peaks is a good measure of signal periodicity.

Applying the above analysis to the temporal dimension of an image sequence, the ratio between the harmonic energy and the total energy of the temporal signal is used here as a measure of the strength of signal periodicity. As a component of the periodicity template of the image sequence, this measure plays an important role in detecting and characterizing periodicity in space and time.

The approach described above assumes that signal periodicity is observable along lines parallel to the temporal (T) axis. In other words, the moving objects need to be tracked just like we fixate on a walking person. Typically, optical flow based techniques are used for object tracking. We present here a non-flow-based frame alignment procedure for tracking, and show that it is more robust to noise than flow based methods.

In this paper, examples of walking people are used to illustrate the technique. However, it should be stressed that the purpose of this work is not to detect and segment a moving object, but to detect and characterize in three-dimensional (3-D) data those regions that exhibit periodicity. We do not expect the algorithm to segment out the walking person. Instead, regions of legs and arms and the outline of the bouncing head and shoulder should be identified.

1.2 Related Work

The work of Polana and Nelson on periodic motion detection [4] is perhaps the most relevant to the approach presented in this paper. In their work, reference curves, which are lines parallel to the trajectory of the motion flow centroid, are extracted and their power spectra computed. The periodicity measure p_f of each reference curve is defined as the normalized difference between the sum of the spectral energy at the highest amplitude frequency and its multiples and the sum of the energy at the frequencies half way between. Besides the value of the periodicity measure itself, there is no checking on the signal harmonicity along the curve, which is a weakness of the method. The periodicity measure for an entire sequence is the maximum of p_f averaged among pixels whose highest power spectrum values appear on the same frequency. The final periodicity measure is used to distinguish periodic and non-periodic motion by thresholding.

In [3], flow based algorithms are used to transform the image sequence so that the object in consideration is stabilized at the center of the image frame. Then flow magnitudes in tessellated frame areas of periodic motion were used as feature vectors for motion classification. In this paper, we show that flow based methods are very sensitive to noise.

This work differ from the above in the following ways: 1) the harmonic relationship among spectral peaks is explicitly verified; 2) a more accurate measure of periodicity in the form of harmonic energy ratios is proposed; 3) multiple fundamentals can be extracted along a temporal line; 4) the values of fundamental frequencies are used in processing to help distinguish periodicity of different activities; 5) regions of periodicity are actually segmented; and 6) the proposed algorithm does not use optical flow, and is robust to noise.

2 Method

The algorithm for periodicity detection and segmentation consists of two stages: (1) object tracking by frame alignment; (2) simultaneous detection and segmentation of regions of periodicity. Object tracking is by itself a research area. Decoupling object tracking and periodicity detection conceptually modularizes the analysis and allows the use of other tracking algorithms.

Throughout this section, an image sequence Walker will be used to illustrate the technical points. More challenging examples are given in Section 3.

2.1 Frame Alignment

In this work, two types of image sequences are considered for frame alignment. In practice, a large number of image sequences can be categorized into one of these two types: (I) area of interest, typically a moving object, is as a whole stationary to the camera, but the background can be moving; (II) little ego-motion is involved and each moving object as a whole is moving approximately frontoparallel to the camera along a straight line and at a constant speed.

Four frames of a sequence with a person walking across the image plane is shown in Figure 1. This is a typical type II sequence. Although there are no re-occurring scenes, we experience the notion of repetitiveness when viewing the sequence. This is due to our ability to fixate on the moving person, so that the person appears to be walking in place. The effect of fixating can be accomplished computationally by realign the image frames. Obviously, frame alignment is not necessary for type I sequences, but in fact is a process of transforming type II sequences into type I.

In the following, the term *data cube* is used to refer to the 3-D (X: horizontal; Y: vertical; and T: temporal) data volume formed by stacking all the frames in a sequence, one

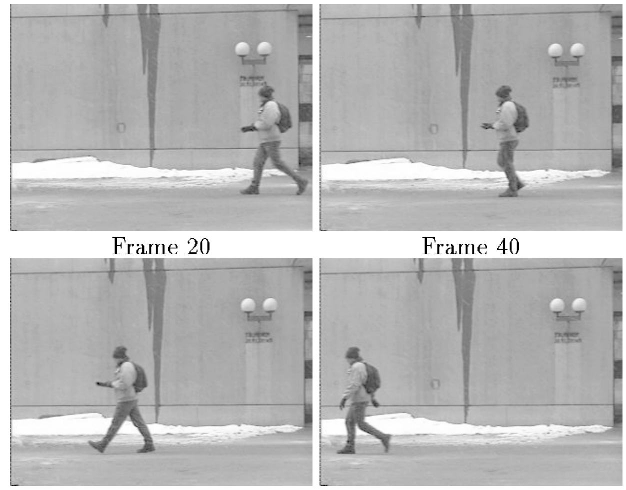
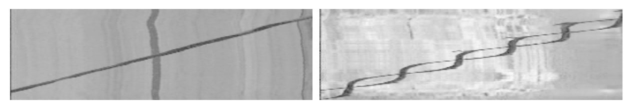


Figure 1: Frames 20, 40, 60, and 80 of the 97 frame Walker sequence, with frame size 320×240.



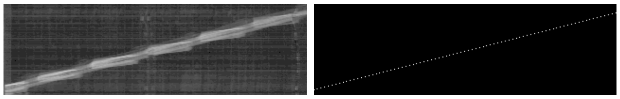
(a) (b)

Figure 2: Head and ankle level XT slices of Walker sequence. (a) Head level. (b) Ankle level. As it is, the periodicity in (b) is difficult to characterize.

in front of the other. The XT and YT slices of the data cube reveal the temporal behavior usually hidden from the viewer. Figure 2 shows the head and ankle level XT slices of the Walker sequence. In (a), the head leaves a nonperiodic straight track while the walking ankles in (b) make a crisscross periodic pattern. As it is, the periodicity in (b) is difficult to characterize. It will be shown that frame alignment transforms data into a form in which periodicity can be easily detected and measured.

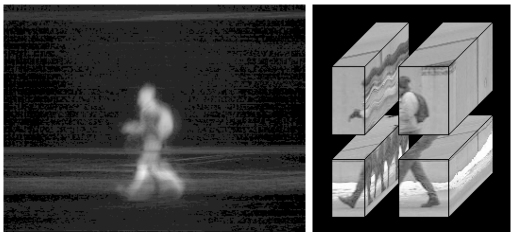
To align a sequence to a particular moving object, the trajectory of the object is first detected. A filtering method similar to the one in [8] is used here to avoid the noise sensitivity of the optical flow based methods (demonstrated in Section 3). Applying 1-D median filtering along the temporal dimension of the sequence (filter length 11 was used for Walker), the resulting sequence has mostly the background. The *difference sequence* between the original and the background contains mainly the moving objects. Since the object trajectories in consideration are approximately linear, the projections of the trajectories onto the XT and YT planes (averaged XT and YT images of the difference sequence) are straight lines. These lines can be detected via a Hough transform to give the X or Y positions of the moving objects in each frame. We call these position values *alignment indices.* The averaged XT image of the Walker difference sequence and the line found by the Hough transform method are shown in Figure 3. Each horizontal line represents a frame, and the diagonal white line marks the object X location in each frame. Note that multiple object trajectories can be detected simultaneously using this procedure, as will be shown in Section 3.1.

Using the alignment indices, image frames in a sequence



(a) (b)

Figure 3: (a) Averaged XT image of the Walker sequence after background removal. (b) Line found in (a) by using a Hough transform method.



(a) (b)

Figure 4: (a) Averaged XY image of aligned Walker difference sequence. The area of interest is clearly shown. (b) Aligned and cropped Walker sequence with splits near the center of the frames to show the inside of the data cube.

can be repositioned to center a moving object to any specified position in the XY plane. After alignment, the object should appear to be moving in place. This in effect is equivalent to fixating on an object when viewing a sequence in which the object's position changes frame by frame. The aligned sequences are passed to the second stage of the algorithm.

2.2 Finding Regions of Periodicity

In the second stage, 1-D Fourier transforms are performed along the temporal dimension of an aligned sequence. The spectral harmonic peaks are detected and used to compute the temporal signal harmonic energy. A periodicity template is generated by using the extracted fundamental frequencies and the ratios between the harmonic energy and the total energy at each frame pixel location. The original sequence is then masked for regions of periodicity.

To save computation and storage, an aligned sequence can be cropped to limit processing to the area of interest. The cropping does not affect the periodicity detection. The location and size of the cropping window can be estimated from the average XY image of the *aligned* difference sequence. Figure 4 shows such XY image of the Walker sequence and the aligned and cropped original sequence with splits near the center of the frames to show the inside of the data cube.

Now consider an aligned and cropped data cube. Frame pixels with the same X and Y locations form straight lines in the cube. Call these lines the *temporal lines*. If the cropped frame size is N_x by N_y, then there are $N_x \times N_y$ temporal lines in the data cube. In the aligned sequence, the object of interest moves in place. If the object is moving cyclically in any manner, the periodicity will be reflected in some of the temporal lines. Figure 5 (a1) and (b1) show the head and the ankle level XT slices of 64 frames (Frame 17 to 80) of the data cube in Figure 4 (b). Each column in the images is a temporal line. These images are the aligned and cropped version of the two XT slices in Figure 2. Columns in Figure 5 (a2) and (b2) are the 1-D power spectra of the corresponding columns in (a1)

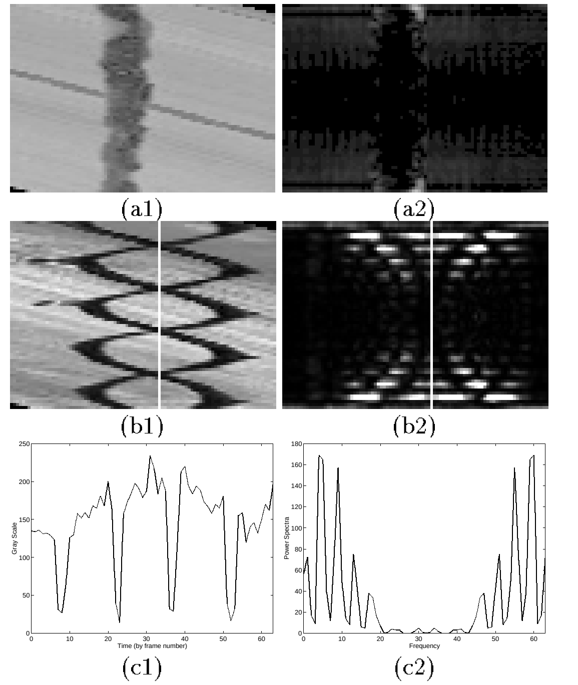


Figure 5: Signals and their power spectra along temporal lines (columns in images). (a1) and (b1): head and ankle level XT slices of aligned and cropped Walker sequence. (a2) and (b2): each column is the 1-D power spectra of the corresponding column in (a1) and (b1). (c1) and (c2): details along the white vertical lines in (b1) and (b2). Periodicity in (b1) is reflected by the spectral harmonic peaks in (b2).

and (b1), normalized among all temporal lines in the data cube. Figure 5 (c1) and (c2) show details along the white vertical lines in (b1) and (b2). While the head level slice in (a1) shows no harmonicity, the periodicity of the moving ankles in (b1) is reflected by the spectral harmonic peaks in (c2). We refer to the spectral energy corresponding to the harmonic peaks as the *temporal harmonic energy* and propose using the *temporal harmonic energy ratio*, which is the ratio between the harmonic energy and the total energy along a temporal line, as a measure of temporal periodicity at the corresponding frame pixel location.

For spectral harmonic peak detection, we adapt the 2-D peak detection algorithm in [6] for 1-D signals. The signal along a temporal line is first zero-meaned and Gaussian tapered, and then its power spectrum computed via a fast Fourier transform. To locate the harmonic peaks, local maxima of the power spectrum are found using size 7 neighborhood and excluding values below 10% of the entire spectral range. A local maximum marks the location of a spectral harmonic peak when its frequency is either a fundamental or a harmonic. A fundamental is defined as a frequency that can be used to linearly express the frequencies of some other local maxima. A harmonic is a frequency that can be expressed as a linear combination of some fundamentals. Starting from the lowest frequency to the highest, each local maxima is checked first for its harmonicity — if its frequency can be expressed as a linear combination of the existing fundamentals, and then for its fundamentality — if the multiples of its frequency, combined with the multiples of existing fundamentals, coincide with the frequency of another local maximum. A tolerance of one sample point is used in the frequency matching. Note that multiple fundamental frequencies can exist along a temporal line.

Due to the nature of the temporal signal and the effect of the Gaussian taper, a spectral harmonic peak usually does not appear as a single impulse. In this work, a peak support region is determined by growing from the detected peak location outward along the frequency axis until the spectral value is below 5% of the spectrum range. After the spectral peaks and their supports are identified, it is straightforward to compute the harmonic energy ratio associated with a fundamental frequency and its harmonics.

The peak detection technique discussed above fails when a temporal line contains only one sinusoidal signal, which produces a single spectral peak. However, this situation arises only when the edge of a moving object has a sinusoidal profile. An example is a vertical sine grating pattern horizontally translating frontoparallel to the camera at a constant speed. Natural edges, patterns, and surfaces hardly ever have such a profile. Therefore, higher harmonics usually accompany the fundamentals of the temporal signals.

Applying the peak detection procedure to all temporal lines in a data cube, the *periodicity template* of the aligned sequence is built by registering the fundamental frequencies and the corresponding values of temporal harmonic energy ratio at each pixel location in a data structure array of frame size. At places where no periodicity is found, the template data structure has value zero. Under circumstances such as a noisy background, some speckles may appear in the template. Simple morphological closing and opening operations can be applied to remove the speckles.

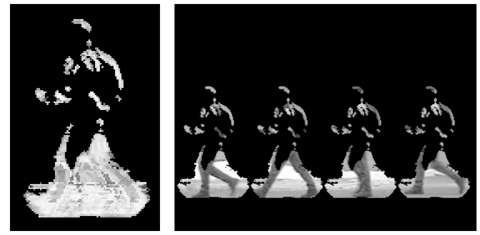
Figure 6 (a) shows the temporal harmonic energy ratio values of the Walker sequence after one closing and one opening operation with a circular structuring element of diameter 3. The larger the energy ratio value, the more periodic energy is at the location. As expected, the brightest region is the wedge shape created by the walking legs. The head, the shoulder, and the outline of the backpack are detected because the walker bounces. The hands appear at the front of the body since in most parts of the sequence the walker was fixing his gloves and moving his hands in a rather periodic manner. Note that the moving background and parts of the walker do not appear in the template since there is no periodicity present in those areas.

Using the alignment indices generated at the first stage, the periodicity template of a sequence can be used to mask the original sequence for the regions of periodicity in each frame. Figure 6 (b) shows the four frames in Figure 1 after they are masked and then stacked together.

Since the non-periodic activities of the background do not light up in the templates, it is clear that the sequence cropping for efficient computation does not affect the processing results.

3 Examples

In addition to the Walker sequence, four examples are used here to demonstrate the effectiveness of the proposed algorithm: Trio, Dog, Wheels, and Jumping Jack.



(a) (b)

Figure 6: (a) Temporal harmonic energy ratio values of the aligned Walker sequence. High value indicates more periodic energy at the location. (b) Using the alignment indices, the four frames in Figure 1 are masked by the template shown in (a) and then stacked together.

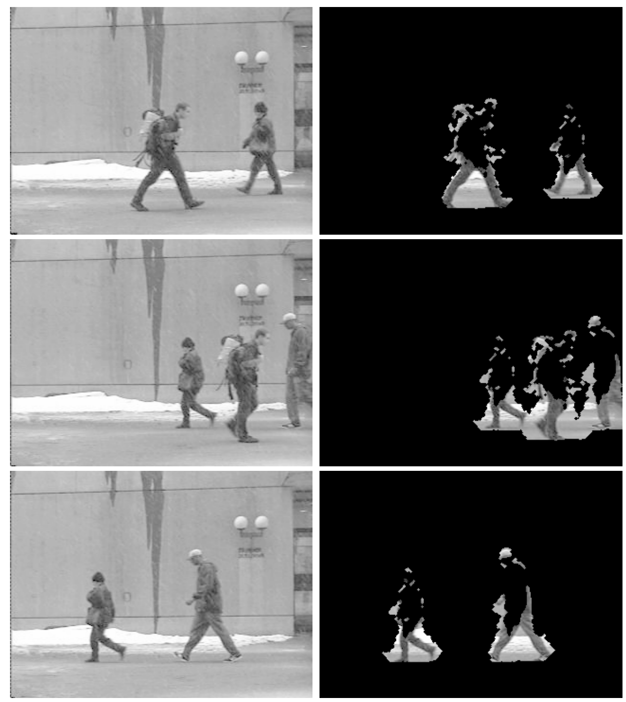
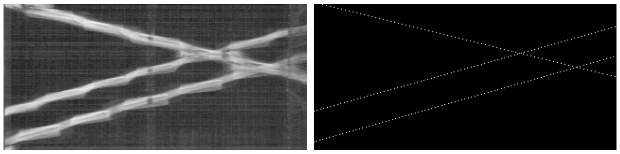


Figure 7: Left column: frames 40, 61, and 88 of the Trio sequence. Right column: frames in the left column masked by the periodicity templates.

The Walker and Trio sequences were recorded by a hand-held consumer-grade camcorder. The Dog and Wheels sequences were taken by the same camera set on a tripod. The Jumping Jack sequence was recorded by a fixed Betacam camera in an indoor setting. Except for the Jumping Jack, none of the subjects in the sequences was aware of the filming; hence the activities are natural and exhibit natural irregularities. All original sequences have 320×240 frame size.

These examples are used to demonstrate 1) the effectiveness of the new algorithm in finding and characterizing periodicity in various settings; 2) the robustness of the algorithm under noisy conditions; and 3) the noise sensitivity of optical flow based estimation methods, which have been used for trajectory detection in many existing works, but are avoided by the method proposed here.



(a) (b)

Figure 8: (a) Averaged XT image of the Trio sequence after background removal. (b) Lines found in (a) by using the Hough transform method.

3.1 Trio

Trio is a 156 frame sequence of three people walking and passing each other. Frames 40, 61, and 88 of the sequence are shown in the left column of Figure 7. As in the Walker example, the averaged XT image is computed after the background removal. The lines in the XT image are detected via Hough transform. Figure 8 shows the averaged XT image and the detected lines. These lines provide the alignment indices of each objects. Note that the alignment indices of three objects are estimated simultaneously.

To generate the periodicity templates, the original sequence is aligned and cropped for each moving person. All aligned sequences contain 64 frames. Figure 9 shows example frames of each aligned sequences and the harmonic energy ratio values of the periodicity templates. Again, the goal here is not to segment out the people, but to detect and characterize regions of periodicity, such as legs, arms, the outline of bouncing head and shoulder, and even the dangling straps of the backpack. Finally, the templates are used to mask the original sequence. Examples are shown in the right column of Figure 7.

Notice that, besides the center person, there is a second or even a third person passing through in all three aligned sequences. However, these passersby have no effect on the results of periodicity detection since they are one-time events on a temporal line, and therefore do not contribute to the temporal harmonic energy. The Trio example demonstrates that the proposed algorithm is well suited for the detection of multiple periodicities, even under the circumstances of temporary object occlusion.

3.2 Dog

Dog is a 104 frame sequence where a person walks two dogs in front of a picket fence. Figure 10 (a) shows frame 46 of the original sequence, and (b) shows frame 13 of the 64-frame aligned sequence. Images (c1) and (c2) show the first and second fundamental frequencies in the periodicity template, while (e1) and (e2) are the corresponding harmonic energy ratios. Note that there are double fundamentals at many pixel locations.

The complication here is the picket fence. In the original sequence, the fence is part of the fixed background, exhibiting pure spatial periodicity. However, when the sequence is aligned to the person and the dogs, the fence starts to move in the background, leaving a periodic signature on many temporal lines. As shown in (c1) and (e1), the fence area lights up in the periodicity template.

Figure 10 (d) shows the fundamentals with value 0.875π, which is the temporal frequency of the fence in the aligned sequence. The fundamental frequency values are used to extract the fence. Figure 10 (f) shows the harmonic energy ratios in the template after the fence frequency is removed. The fence region of the frame in (a) is shown in (g) while other regions of periodicity are shown in (h).

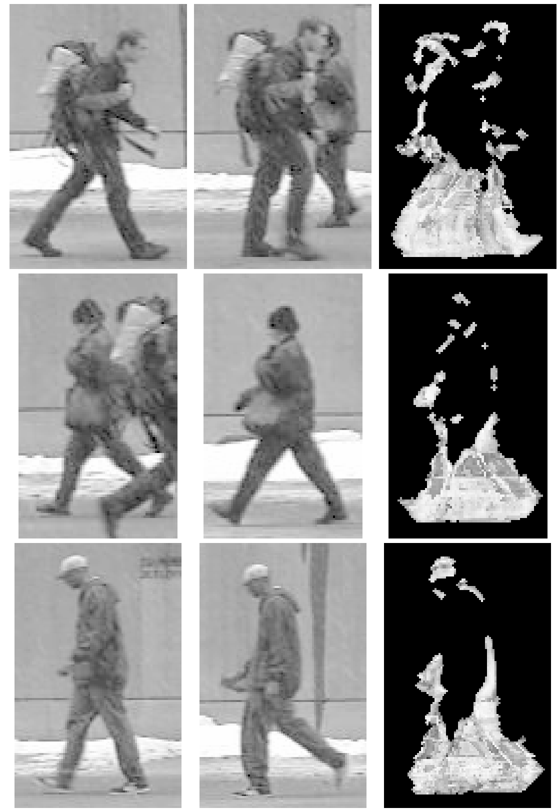


Figure 9: Example frames of aligned sequences and the harmonic energy ratio values of the periodicity templates for each individuals of the Trio sequence. First two columns: example frames. Right column: harmonic energy ratio values.

3.3 Wheels

The examples shown so far all involve walking. However, the algorithm is not limited to periodicity caused by human activities, but works in general for any periodic space-time phenomenon.

Wheels is a 64 frame sequence of a car passing by a building. Near the top of the building, two spinning wheels are connected by a figure 8 belt. One side of the belt is patterned and appears periodic. Every region with periodicity should be captured: the hub caps, the wheels, and one side of the belt. As shown in Figure 11, the algorithm accomplishes just that.

3.4 Jumping Jack

There is no translatory motion in the Jumping Jack sequence, and the background is smooth. This sequence and its noisy versions (corrupted by additive Gaussian white noise (AGWN) of variance 100 and 400) are used to demonstrate the robustness of the new algorithm in the presence of noise, and also to show the noise sensitivity of the optical flow based motion estimation. The length of the sequences used here is 128 due to the cycle of the jumping motion.

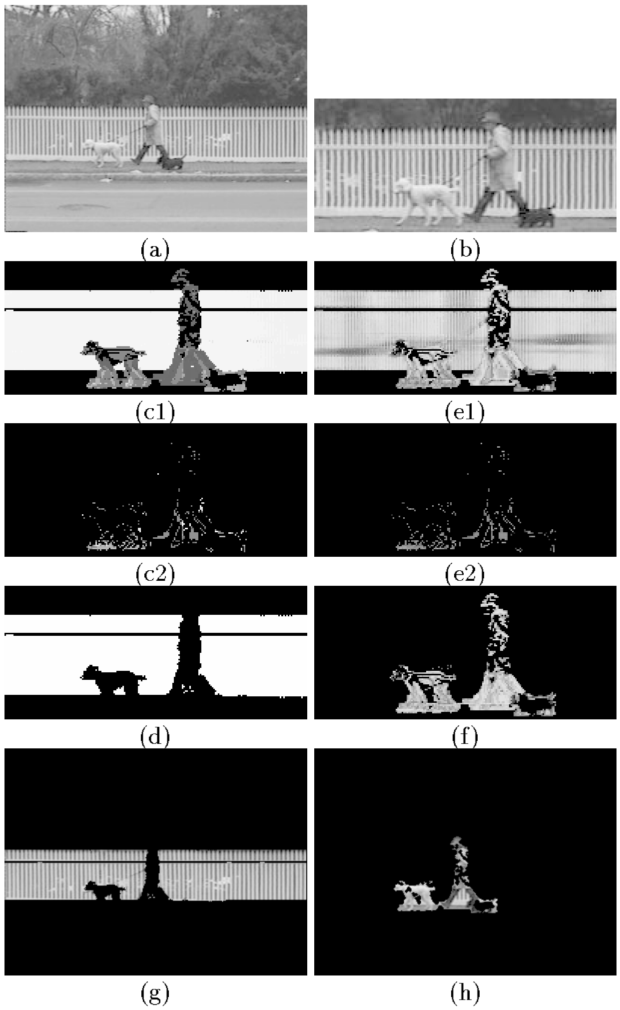


Figure 10: Dog sequence. (a) Frame 46 of original sequence. (b) Frame 13 of aligned sequence. (c1) and (c2): first and second fundamental frequencies in periodicity template. (e1) and (e2): harmonic energy ratios corresponding to the frequencies in (c1) and (c2). (d) Fundamentals with the fence frequency. (f) Harmonic energy ratios after the fence frequency is removed. (g) Frame 46 masked to show fence region. (h) Frame 46 masked to show other regions of periodicity.

Most of the related work uses flow based methods to locate moving objects in a sequence. However, the noise sensitivity of the flow based method can be a drawback. The optical flow magnitudes shown here were obtained by using the hierarchical least-squares algorithm [9], which is based on a gradient approach described by [10] [11]. Two pyramids are built, one for each of the two consecutive frames, and motion parameters are progressively refined by residual motion estimation from coarse images to the higher resolution images. This algorithm is representative of the existing optical flow estimation techniques. The optical flow magnitudes of the Jumping Jack frame 61 are shown in the second row of Figure 12. Given a clean input, the flow magnitudes can be used to segment the moving object. However, the algorithm is mostly ineffective under the noisy conditions.

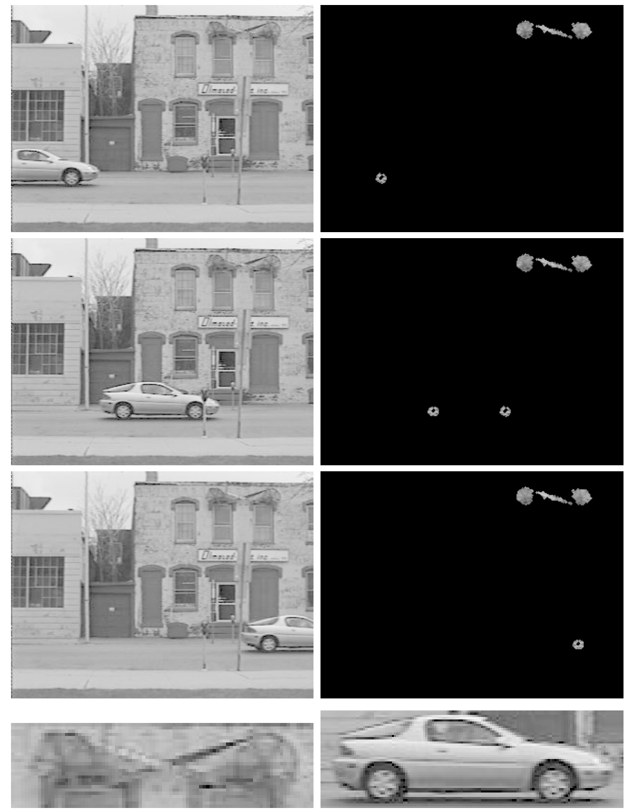


Figure 11: Wheel sequence. Shown in the first three rows, the algorithm captures all regions with periodicity: the hub caps, the wheels, and one side of the belt. Bottom row: details of spinning wheels and car.

The third row of Figure 12 shows the 57th TY (not YT!) image of each sequence, revealing the tracks left by the right hand and leg. The rows in these images are temporal lines, and the corresponding power spectra are shown in the fourth row of the figure. The periodicity templates can be found in the bottom row. Although the noise causes some degradation in the arm regions, the templates are well preserved overall. The reason why the proposed algorithm is not affected by large amounts of white noise in the input is that white noise only contributes to the relatively smooth part of the power spectrum. As long as the noise energy is not so high that it overwhelms the spectral harmonic peaks, the algorithm works.

3.5 Walker

The detection results of the Walker sequence were shown in Section 2. Here we show the results from noisy inputs (original sequence corrupted by AGWN of variance 100 and 400), using 64 frames. The resulting periodicity templates in Figure 13 show that, unlike optical flow based methods, the proposed algorithm is robust in the presence of noise.

4 Discussion

Compared to the method used in [4], the periodicity measure proposed here in the form of the temporal harmonic energy ratio is a more accurate and more reliable measure

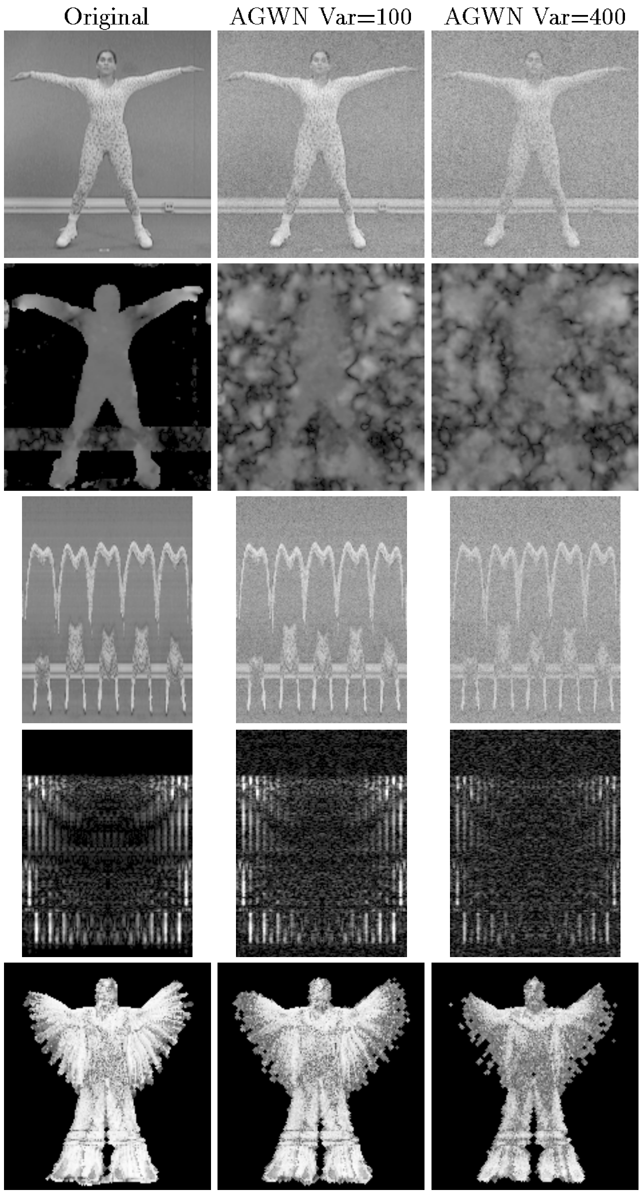


Figure 12: Jumping Jack sequence (frame size 155×170). Left column: original sequence. Middle and right columns: corrupted sequences. Row 1: frame 61. Row 2: frame 61 optical flow magnitudes. Row 3: TY slice 57, showing the tracks left by right hand and leg; each row of these images is a temporal line. Row 4: temporal power spectra of TY slice 57. Row 5: harmonic energy ratios of periodicity templates.

of signal periodicity.

The periodicity templates also provide the fundamental frequencies of the temporal signals. Using this information, areas involved in periodic activities with different cycles can be distinguished easily.

The proposed algorithm can be considered as a "periodicity filter". Given a sequence of a street with cars and pedestrians, the algorithm will find the moving legs of the pedestrians and filter out the cars and other non-periodic

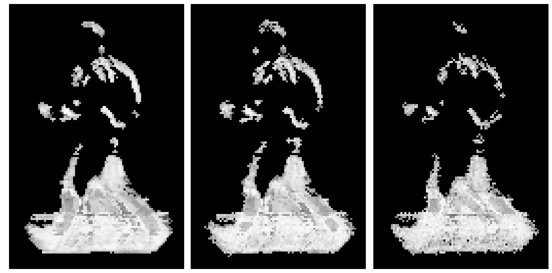


Figure 13: Periodicity templates of the Walker sequence. (a) from original sequence; (b) and (c): with AGWN of variance 100 and 400 respectively. The proposed algorithm is robust in the presence of noise.

activities. Periodicity is a salient feature to human visual perception. The proposed algorithm provides a model of low-level periodicity perception, even though it may not work exactly like the human visual system.

The method presented here is computationally efficient. The most machine intensive part of the algorithm is the 1-D fast Fourier transform used in power spectrum computation. When the activity cycle is reasonably short, such as walking in normal speed, a sequence length of 64 frames suffices. Cropping of aligned sequences provides additional speed-up.

In the current work, assumptions were made on the data. The steady background condition for data type II is mainly for the background subtraction. The algorithm in fact tolerates small camera movement quite well. When an object is not translating with respect to the camera, its trajectory will not be linear in the data cube and a scheme more sophisticated than the Hough line detection will have to be used for the frame alignment. If the object is not moving frontoparallel to the camera, the perspective effect will change the size of the object in the sequence. However, this change should not be significant during the period of 64 frames when the distance between the camera and the object is sufficiently large. In practical situations, this is often the case.

4.1 Applications

The proposed algorithm can be applied to motion classification and recognition. In [5], the shape of the active region in a sequence was used for activity recognition. In [3], the sum of the flow magnitudes in tessellated frame areas of periodic motion was used for motion classification. The periodicity templates produced by the proposed algorithm can provide not only distinct shapes of regions of periodic motion, such as the wedge for the walking motion and the snow angle for the jumping jack, but also accurate pixel-level description of a periodic action in the form of temporal harmonic energy ratios and motion fundamental frequencies.

The characterization of periodicity is also important to video database related applications. The presence, position, strength, and frequency information of periodic activities can be used for video representation and retrieval.

In general, periodicity is a salient attention-getting feature. The proposed algorithm can be used in numerous surveillance applications for detecting ambulatory activity without having to do full-person recognition.

5 Summary

A new algorithm for finding periodicity in space and time is presented. The algorithm consists of two main parts: 1) object tracking by frame alignment, which transforms data into a form in which periodicity can be easily detected and measured; 2) Fourier spectral harmonic peak detection and energy computation to identify regions of periodicity and measure its strength. This method allows simultaneous detection, segmentation, and characterization of spatiotemporal periodicity, and is computationally efficient. The effectiveness of the technique and its robustness to noise over optical flow based methods are demonstrated using a variety of real-world video examples.

Periodicity templates are proposed as a new way of characterizing spatiotemporal periodicity. The templates contain information such as the fundamental frequencies and the temporal harmonic energy ratios at each frame pixel location. The periodicity templates and the template generating algorithm are useful tools for applications such as action recognition, video databases, and video surveillance.

Acknowledgments

The authors would like to thank Aaron Bobick and Sandy Pentland for insightful discussions, Jim Davis for the Jumping Jack sequence, and John Wang for the hierarchical optical flow estimation program.

References

[1] D.D. Hoffman and B.E. Flinchbuagh. The interpretation of biological motion. *Biological Cybernatics.*, pages 195–204, 1982.

[2] M. Allmen and C.R. Dyer. Cyclic motion detection using spatiotemporal surface and curves. In *Proc. ICPR*, pages 365–370, 1990.

[3] R. Polana and R. Nelson. Low level recognition of human motion. In *IEEE Workshop on Motion of Non-rigid and Articulated Objects*, pages 77–82, Austin, TX, Nov. 11-12 1994.

[4] R. Polana and R. C. Nelson. Detecting activities. In *Proc. CVPR*, pages 2–7, New York, NY, June 1993.

[5] A.F. Bobick and J.W. Davis. Real-time recognition of activity using temporal templates. In *Proc. Third IEEE Workshop on Appl. of Comp. Vis.*, pages 39–42, Sarasota, FL, Dec. 1996.

[6] F. Liu and R. W. Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE T. Pat. Analy. and Machine Intel.*, 18(7):722–733, July 1996.

[7] H. Wold. *A Study in the Analysis of Stationary Time Series.* Stockholm, Almqvist & Wiksell, 1954.

[8] S. A. Niyogi and E. H. Adelson. Analyzing gait with spatiotemporal surfaces. In *IEEE Workshop on Motion of Non-rigid and Articulated Objects*, pages 64–69, Austin, Texas, Nov. 11-12 1994.

[9] J.Y.A. Wang. *Layered Image Representation: Identification of Coherent Components in Image Sequences.* PhD thesis, Dept. of EECS, MIT, Cambridge, Sept. 1996.

[10] J. Bergen *et al.* Hierarchial model-based motion estimation. In *Proc. ECCV*, pages 237–252, 1992.

[11] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Image Understanding Workshop*, pages 121–131, 1981.