M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 400

Detecting and Segmenting Periodic Motion *

Fang Liu and Rosalind W. Picard
The Media Laboratory, E15-390
Massachusetts Institute of Technology, Cambridge, MA 02139
fliu@media.mit.edu, picard@media.mit.edu

Abstract

A new algorithm is presented for the detection and segmentation of multiple periodic motions in image sequences. A periodicity measure which provides an accurate quantitative measure of signal periodicity in the form of temporal harmonic energy ratio is proposed. The periodicity templates which incorporate the segmented region with the associated periodicity energy distribution and motion fundamental frequencies provide a new means of periodic motion representation. The algorithm is computationally simple and robust in the presence of noise. Real world examples are used to illustrate the techniques.

1 Introduction

1.1 Overview

Periodic motion is common in the natural world. For instance, locomotion often exhibits cyclic movement. Periodicity is also an important cue in human motion perception. Information regarding the nature of a periodic motion, such as location, extent, and frequency, is important for motion understanding. Techniques for periodic motion detection and segmentation can assist in many applications requiring object and activity recognition and representation.

The main body of work on periodic motion is model-based (eg. [1][2]). More recently there is work on motion recognition directly using low-level features of motion information (eg. [3][4][5]). However, to date, there has not been a method which uses low-level motion features to detect and segment periodic motion simultaneously. In this work, we attempt to tackle this problem by using a Fourier spectral based approach. The periodicity templates generated in the process incorporate the location, extent, and other characteristic information, such as frequency, of a periodic motion. Therefore, the templates can also be used for periodic motion representation.

1.2 Our Approach

The term *temporal texture* is defined in [4] as "the motion patterns of indeterminate spatial and temporal extent". We would like to use "texture" to refer to any approximately homogeneous phenomenon. In this sense, periodic temporal activity is a type of temporal texture.

The algorithm we present here is motivated by theory for textured static images that assumes an underlying random field representation for the data [4]. In particular, 1-D signals along the temporal dimension can be considered as stochastic processes. When assuming stationarity, a stochastic signal can be decomposed into deterministic (periodic) and indeterministic (random) components (Wold decomposition [6]). In the frequency domain, the two components correspond to the singular and continuous part of the Fourier spectra respectively. 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 spectra. Therefore, the energy contained in the harmonic peaks is a good measure of periodicity in the signal. In this work, we apply this analysis to the temporal dimension of the image sequences and propose a new measure of temporal periodicity. This measure is then used to build a periodicity template for simultaneous periodic motion detection and segmentation.

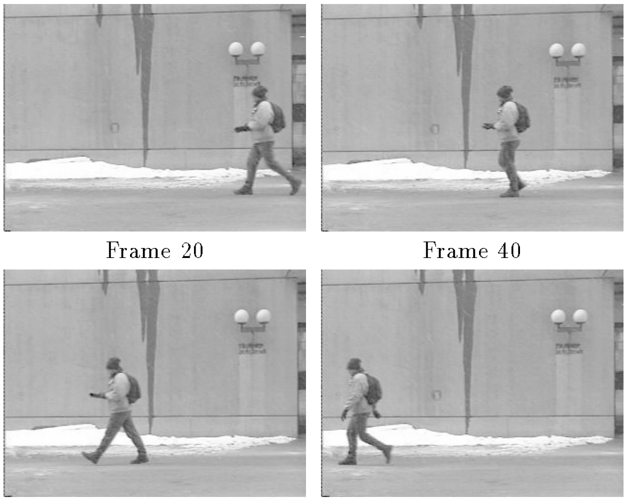
The actual implementation of the approach above assumes that the repetitiveness of an action is observable along lines parallel to the temporal (T) axis. In other words, the moving object needs to be tracked just like we fix our eyes on a walking person. Typically, optical flow based techniques are used for object tracking. However, flow based methods are usually susceptible to noise. We present here a non-flow-based procedure for foveating on moving objects by frame alignment.

1.3 Related Work

Of the large amount of research in this area, 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

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



Frame 20 Frame 40

Frame 60 Frame 80

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

image sequence so that the object in consideration stays 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. There is no precise segmentation of regions of periodic motion involved.

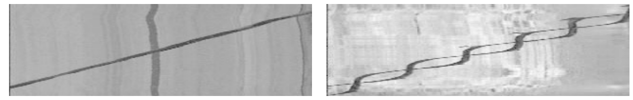
2 Method

2.1 Overview

Our system for periodic motion detection and segmentation consists of two stages: 1) foveating by frame alignment; 2) simultaneous detection and segmentation of regions of periodic motion.

In this work, two types of image sequences are considered. (1) Each object which may have a periodic motion is as a whole stationary to the camera, but the background can be moving. (2) The relative motion between the camera and the background is small, permitting minor shake and gradual drift as in the hand-held situation, and each moving object is as a whole moving approximately frontoparallel to the camera at relatively constant speeds and in a translatory manner. In practice, large number of image sequences containing natural periodic motions can be categorized into one of the two types.

When watching a sequence of a person walking across the image plane, we experience a notion of repetitive movement. However, if examining the individual frames, there would be no re-occurring scenes. Several frames of a sequence Walker are shown in Figure 1 (see also http://[web-seq-1]). The reason why we see a periodic motion in the sequence is due to our ability to focus our visual attention on the moving object, or, *foveat*. The effect of foveating can be accomplished computationally by frame alignment. Obviously, foveating is not necessary for sequence type (1), but in fact is a process of transforming sequences of type (2) into type (1).



(a) (b)

Figure 2: Head and ankle level XT slices of the Walker sequence. (a) Head leaves a straight track. (b) Walking ankles make a crisscross pattern

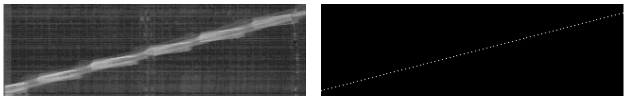
In the second stage, 1-D Fourier transforms are performed on each pixel location along the temporal dimension of the aligned frames. The spectral harmonic peaks are then detected and used to compute harmonic energy involved in each frame pixel location. A periodicity template in frame size is generated based on the ratio between the harmonic energy and the total energy. The original sequence is then masked for regions of periodic motion.

In the following, we use the term *data cube* to refer to the three dimensional (X: horizontal; Y: vertical; and T: temporal) data volume formed by putting together all the frames in a sequence, one behind 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. The head leaves a more or less straight track while the walking ankles make a crisscross pattern. Throughout this section, the Walker sequence will be used to illustrate the technical points. More examples are given in Section 3.

2.2 Foveating by Frame Alignment

To align a sequence to a particular moving object, we need first to detect the trajectory of the object. In existing works, certain kinds of tracking techniques based on differencing consecutive frames are usually used to locate the moving objects. The flow-based methods are inevitably subject to the noise sensitivity (which will be demonstrated later) inherent in pairwise frame comparison. To avoid these shortcomings, we take advantage of the two types of restrictions on data and use a method similar to the one in [7] to look directly in the data cube for tracks left by the moving objects.

Since the background of the sequence changes slowly, a 1-D median filter can be applied to the data volume along the temporal dimension T to exclude moving objects and result in a sequence containing mostly the background. Filter length of 11 was used in the Walker sequence. Subtracting the background from the original sequence, we obtain a *difference sequence* containing mainly the moving objects. By the condition (2) on data, the trajectories of the objects as a whole in the difference cube should be approximately linear. To obtain 2-D representations of the trajectories, we simply compute the average of the XT or the YT slices of the difference cube, which is equivalent to collapsing the difference cube top down to the XT plane or sideways to the YT plane. Next, lines in the 2-D trajectory images are detected via the Hough transform. The detected lines give the X or Y positions of the moving objects in each frame. These position values are called *alignment indices*. The averaged XT image of the Walker



(a) (b)

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

difference sequence and the line found by the Hough transform method are shown in Figure 3. Note that multiple object trajectories can be detected simultaneously using this procedure. An example of three walking persons is shown in Section 3.

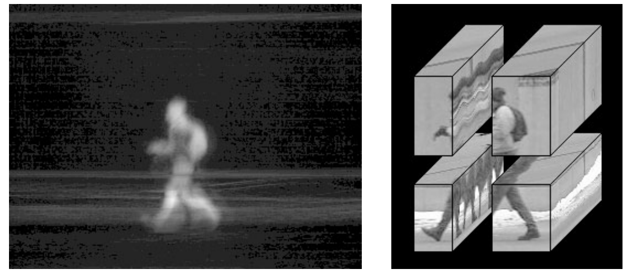
Using the alignment indices of an object, each frame in the image sequence can be repositioned to center the object to any specified position in the XY plane. After alignment, the object should appear as moving in place. For instance, after aligning the Walker sequence in the X dimension (alignment in Y dimension is not necessary since the overall Y position of the person does not change much), the position of the walker's torso remains in place, but the surroundings move to the right. In effect, this is equivalent to focusing our visual attention on an object when viewing a sequence in which the object's position changes frame by frame. We call this process *foveating by frame alignment*. The aligned sequences are passed on to the second stage of the system.

2.3 Detection and Segmentation

In this stage, we only deal with sequences already aligned to a particular object. To save computation and storage, an aligned sequence can be cropped to limit processing to the object and its vicinity. It will become clear later that the cropping has no effect on the estimation of periodic motion. The location and size of the cropping window is estimated from the average XY image of the difference sequence which is first *aligned* to the particular object. Figure 4 shows the averaged aligned XY image and the aligned and cropped Walker sequence with splits near the center of the frames to show the inside of the data cube.

Facing the data cube, a line can be drawn from a pixel (x_0, y_0) in the first frame all the way through the cube to arrive at pixel (x_0, y_0) in the last frame. Clearly, this line contains pixel (x_0, y_0) of all frames. We name this line the *temporal line* at (x_0, y_0), which is of essential importance to our discussion. If the frame size is N_x by N_y, then there are $N_x \times N_y$ temporal lines in the data cube.

Since the image sequence is aligned to a particular object, the object should be moving in place. Apparently, if the object is moving cyclically in any manner, the periodicity will be reflected in some of the temporal lines. The top two images of Figure 5 shows the head and ankle level XT slices of 64 frames (Frame 17 to 80) of the data cube in Figure 4 (b). These slices are in fact the aligned and cropped version of the two XT slices in Figure 2. Every column in the two images is a temporal line. The Fourier power spectra of these temporal lines are the columns of the corresponding bottom two images in Figure 5. Note



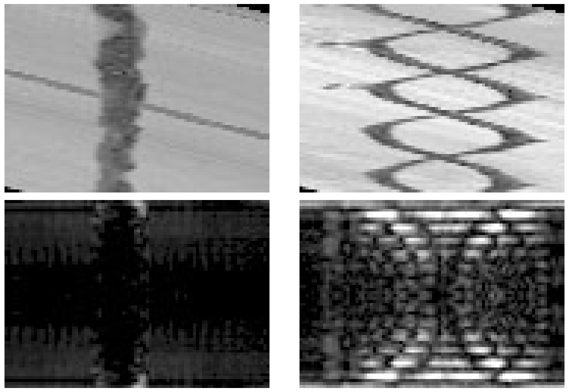
(a) (b)

Figure 4: (a) Averaged XY image of the aligned Walker sequence after background removal. (b) Aligned and cropped Walker sequence with splits near the center of the frames to show the inside of the data cube.

that the power spectrum values are normalized among all temporal lines in the data cube. While the spectra of the head slice shows no harmonic peaks, the periodicity of the moving ankles is captured by the spectral harmonic peaks. We refer to the spectral energy corresponding to these harmonic peaks as the *temporal harmonic energy* and propose using the ratio between the harmonic energy and the total energy of a temporal line (*temporal harmonic energy ratio*) as a measure of temporal periodicity at the corresponding pixel location.

In [8], a method was presented to detect spectral harmonic peaks in 2-D random fields. Here the algorithm is adapted to use for 1-D signals. Given a temporal line, it is first zero-meaned and Gaussian tapered, then its power spectrum values are computed via a fast Fourier transform. To locate the harmonic peaks, local maxima of the spectrum values (excluding values below 10% of the value range in the data cube) are found by searching a size 7 neighborhood of each frequency sample. These local maxima provide candidate locations of harmonic peaks. A local maximum marks the location of a harmonic peak only when its spectral frequency is either a fundamental or a harmonic. A fundamental is defined as a frequency which can be used to linearly express the frequencies of some other local maxima. A harmonic is a frequency which 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.

Due to the nature of the natural temporal signal and the windowing effect in spectra computation, a harmonic peak usually does not appear as a single impulse. Therefore the procedure so far only provides the summit location of each peak. The support of a peak is determined by growing outward from the summit location along the frequency axis until the spectrum value is below certain small value (5% of the spectrum value range in this work). After the harmonic peaks are identified, it is straightforward to compute the



(a) (b)

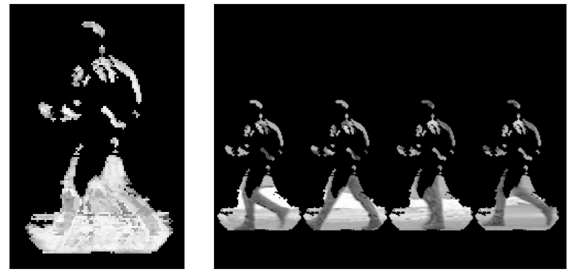
Figure 5: Temporal lines and their normalized power spectra at the head and ankle level of the aligned Walker data cube. Top row: XT slices, every column in the images is a temporal line. Bottom row: power spectra, each column is the power spectra of the corresponding column in the XT slice. (a) Head level, showing no harmonic peaks. (b) Ankle level, periodicity of the moving ankles is captured by the harmonic peaks in spectra.

harmonic energy ratio of the temporal line.

One may argue that the technique discussed above will fail when a temporal line contains only a sinusoidal signal which produces only a single spectral peak. Theoretically, this is correct. However, this situation almost never arises in the image sequences of natural scenes and objects. To explain this, we trace back to the formation of the signal on a temporal line. Since a temporal line corresponds to a particular pixel in the image plane, having a pattern moving across the pixel is equivalent to having the pixel scanning across the pattern. When will this scan create a sinusoidal signal? The answer is only when the pattern has a sinusoidal profile. (An example is to translate horizontally a vertical sine grating pattern frontoparallel to the camara in a constant speed.) However, natural edges, patterns, and surfaces hardly ever have such a profile. Therefore, it is safe to say that higher harmonics will usually accompany the fundamentals in the Fourier Spectra of the temporal lines.

Applying the peak detection procedure to all temporal lines in a data cube and registering the temporal harmonic energy ratio at each pixel location as gray level values in a blank plane of frame size, a *periodicity template* for periodic motion is built for the aligned sequence. The larger the template value, the more periodic energy at the location. At places where no periodic motion is involved, the template value remains zero. Under circumstances such as a noisy background, some speckle noise may appear in the template. Simple morphological closing and opening operations can be applied to the template to clean up the speckle.

Figure 6 (a) is the periodicity template of the Walker sequence after one closing and one opening operation with a 3 pixel diameter circular structuring element. 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 shown 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 periodic motion in those areas.



(a) (b)

Figure 6: (a) Periodicity template generated from the aligned Walker sequence of Figure 4 (b). The value at each pixel is that of the temporal harmonic energy ratio of the corresponding temporal line. High value indicates more periodic energy at the location. (b) Using the periodicity template to mask the original sequence. The four frames in Figure 1 are masked and stacked together into one frame.

Since the non-periodic motion of the background does not light up in the templates, it is clear that the sequence cropping earlier in the second stage does not effect the algorithm itself, but only increases the computational efficiency.

Using the alignment indices generated at the first stage, the periodicity template can also be applied to the original sequence to mask the regions of periodic motion in each frame. The masked frames corresponding to the ones in Figure 1 are stacked together and shown in Figure 6 (b) (see also http://[web-seq-2]).

3 Examples

Three example sequences are used in this paper. One is the Walker sequence. Another walking sequence is Trio, where three people walking and passing each other. The last one is an aerobics sequence called Jumping Jack. Walker and Trio are taken by a hand-held consumer-grade camcorder during a snow storm. Camera drift and the influence of breathing of the cameraman are visible in the sequences. Jumping jack is taken by a fixed Betacam camera in an indoor setting. These examples are used to demonstrate 1) the effectiveness of the new algorithm in the detection and segmentation of both single and multiple objects with periodic motions; 2) the robustness of the algorithm under noisy conditions; 3) the noise sensitivity of optical flow based estimation methods, which are used for trajectory detection in many existing works, and avoided by the method proposed here.

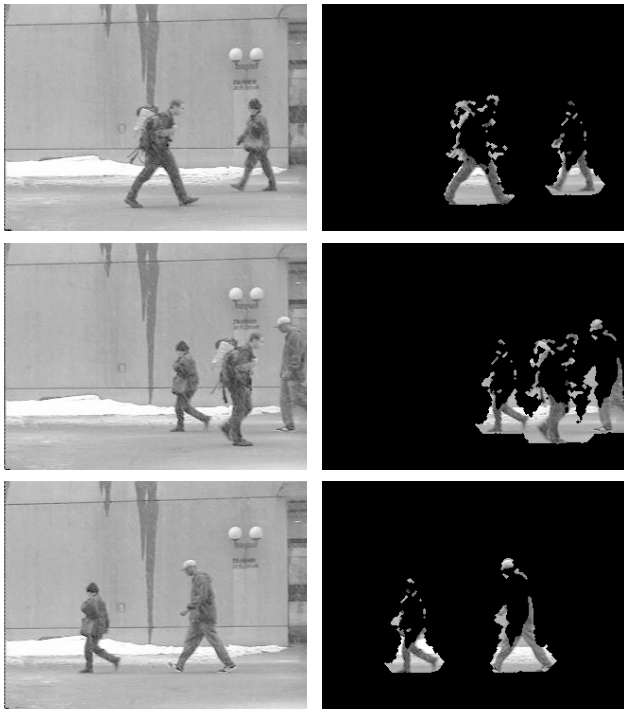


Figure 7: Left column: frames 40, 61, and 88 of the 156 frame sequence Trio, with frame size 320 by 240. Right column: frames 40, 61, and 88 masked by the periodicity templates of three individuals.

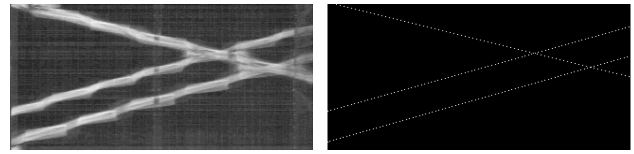
3.1 Trio

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 (see also http://[web-seq-3]).

As in the Walker example, a temporal median filter is used to extract the background. After the background is largely removed from the sequence, the averaged XT image is computed. The lines in the XT image are then detected via Hough transform. Figure 8 shows the averaged XT image and the lines detected from the image. The detected lines provide the alignment indices of each objects. Note that the alignment indices of all three objects are estimated simultaneously.

Next, as in the Walker example, the original sequence is aligned and cropped for each of the three moving individuals. All aligned sequences contain 64 frames. Then the aligned sequences go though the process of power spectrum estimation and harmonic peak detection. Finally, the temporal harmonic energy ratio at each pixel location is computed to generate the periodicity templates. Figure 9 shows example frames of each aligned sequences and the periodicity templates. The templates are then used to mask the original sequence. Examples of masked sequence are shown in the right column of Figure 7 (see also http://[web-seq-4]).

Notice that, besides the center person, there is a second



(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.

or even a third person passing through in all three aligned sequences. However, their appearance has no effect on the result of periodic motion detection and segmentation. This is because, to any individual temporal line, these passersby are one-time event and do not contribute to the temporal harmonic energy of the temporal line. The Trio example demonstrates that the proposed algorithm is well suited for the detection and segmentation of multiple periodic motions, even under the circumstances of temporary object occlusion.

3.2 Jumping Jack

In the Jumping Jack sequence, there is no translatory motion involved and most part of the background is very smooth. We use this sequence and the noisy versions of it to demonstrate the robustness of the new algorithm under noisy conditions and also to show the sensitivity of the optical flow based motion estimation to noise. There are three different kinds of input given to the system: the original sequence and the sequences corrupted by additive Gaussian white noise (AGWN) with variance 100 and 400. The length of the sequences used in power spectrum estimation is increased to 128 due to the cycle of the jumping motion. All the data related to the Jumping Jack sequence are shown in Figure 10 (see also http://[web-seq-5] to [web-seq-10]).

The second row in Figure 10 are the 57th TY (not YT!) image of each data cube. They show 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 third row of the figure. The periodicity templates can be found in the fourth row of the figure. Although the noise does cause some degradation in the arm region, other areas of the templates are well preserved.

The reason why the proposed algorithm is not affected by certain amount of white noise in the input is that white noise only contributes to the continuous component of the power spectrum. As long as the noise energy is not so high that it overwhelms the spectral harmonic peaks, the algorithm works.

Most of the related works use flow based methods to extract spatiotemporal surfaces or curves to locate the moving objects in a sequence, but we would like to demonstrate here that the noise sensitivity of the flow based method can be a drawback in real applications. The optical flow magnitudes are estimated here by using the hierarchical least-squares algorithm [9] which is based on a gradient approach described by [10] [11]. Two pyramids are built,

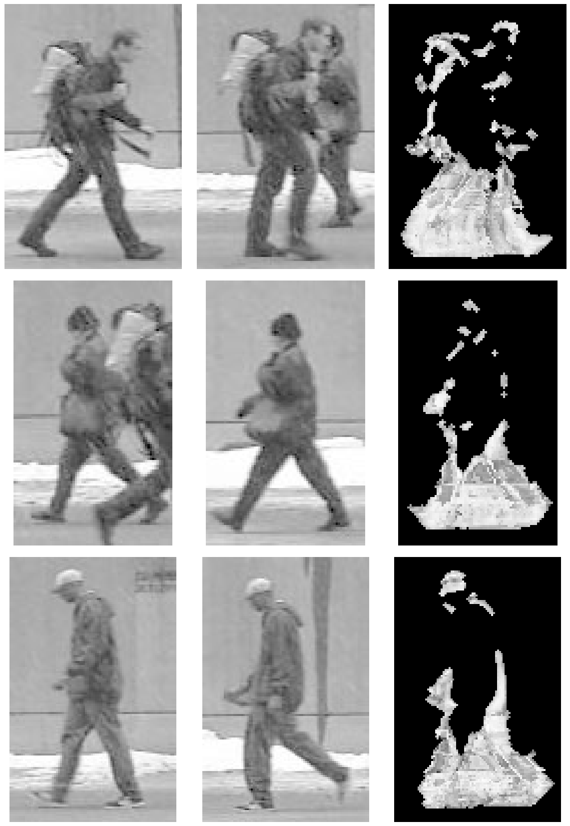


Figure 9: Example frames of aligned sequences and the corresponding periodicity templates for each individuals of the Trio sequence. Two left columns: example frames. Right column: periodicity templates.

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 last row of images in Figure 10 is the optical flow magnitudes of frame 61 in the sequences. When given a clean input such as the original Jumping Jack sequence, the flow magnitudes can be used to segment out the moving object. However, under the noisy condition, the algorithm is mostly ineffective.

3.3 Walker

The detection and segmentation results of the Walker sequence are mostly shown in the previous section. Here we show the results from noisy inputs. Additive Gaussian white noise of variance 100 and 400 was used to corrupt the original sequence. The length of the aligned sequences for power spectrum estimation is 64. The resulting periodicity templates in Figure 11 clearly show that the proposed algorithm is robust in the presence of noise.

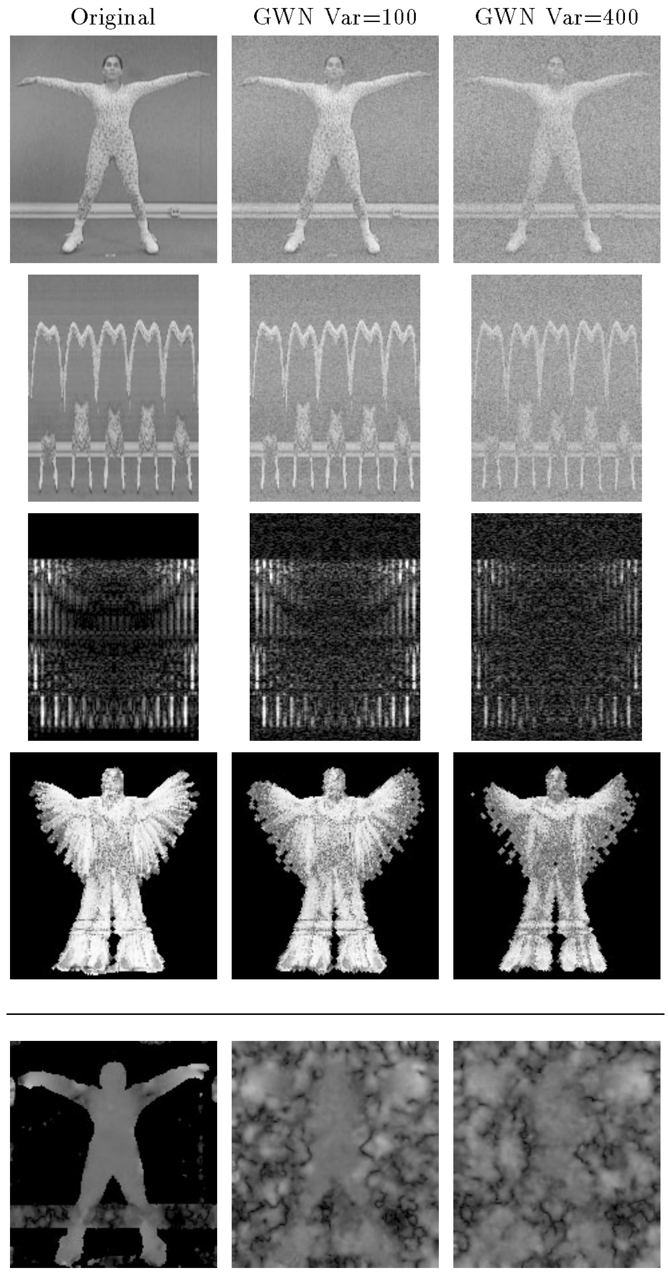
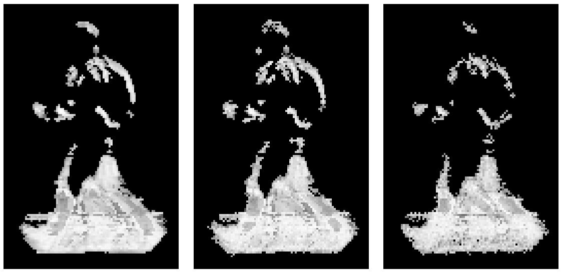


Figure 10: Data of 128 frame Jumping Jack sequence with frame size 155 by 170. Left column: original sequence. Middle column: corrupted by AGWN with variance 100. Right column: corrupted by AGWN with variance 400. Row 1: frame 61 of Jumping Jack sequence. Row 2: TY slice 57, showing the tracks left by the right hand and leg; each row of these images is a temporal line. Row 3: temporal line power spectra of TY slice 57. Row 4: periodicity templates. Row 5: optical flow magnitudes of frame 61.



(a) (b) (c)

Figure 11: Periodicity templates of the Walker sequence: (a) from original sequence; (b) with AGWN of variance 100; (c) with AGWN of variance 400. The proposed algorithm is robust in the presence of noise.

4 Discussion

4.1 Algorithm

Compared to the one used in [4], the periodicity measure proposed here in the form of the temporal harmonic energy ratio is a more accurate and reliable measure of signal periodicity. It not only can indicate the presence of the periodic behavior, but also gives a quantitative measure of how much energy at each pixel location is contributed by the periodicity.

The use of the periodicity template allows the detection and segmentation of the periodic motion to be accomplished simultaneously. Since the fundamental frequencies of the temporal lines are extracted in the process of computing the templates, they can be registered to the templates as well. Using this information, areas involved in periodic motion with different cycles can be easily distinguished. This is useful in situations such as a person walking in front of a picket fence. When the sequence is aligned to the person, the fence will be moving in the background and leave a periodic signature on some temporal lines. Consequently, the fence area will light up in the periodicity template. However, the fundamental frequency of the moving fence is usually quite different from the one of the walker.

The proposed algorithm can also be considered as a periodic motion filter. At first, all moving objects are targets for foveating, but then only the ones exhibiting periodic motions remain. Consider an input of a street with cars and pedestrians, the system would filter out the cars and keep the pedestrians.

Existing related work often uses flow based methods to extract the trajectories of moving objects. Since flow based methods can be susceptible to noise, we instead use foveating by frame alignment to focus on individual objects. This approach not only helps to improve the performance of the system in the presence of noise, but also is efficient in that it generates alignment parameters of all moving objects at once.

The method introduced here is not computationally taxing. The most machine intensive part of the algorithm is the 1-D fast Fourier transform used in power spectrum computation. However, when the motion cycle is reasonably short, such as walking in normal speed, sequence length of 64 suffices. Cropping of aligned sequences also helps to speed up the processing.

In the current work, we put restrictive conditions on data. The steady background condition in the translatory moving object case is for the background subtraction. The system actually tolerates small camera movement quite well. When an object is not moving in a translatory manner 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 large during the period of 64 frames when the distance between the camera and the object is sufficiently large comparing to the size of the camera, and in many practical situations, this is the case. When the data is reasonably clean, flow based tracking and scaling methods such as the one in [3] can also be used for frame alignment.

4.2 Applications

Among other possible applications, the proposed algorithm can be applied readily 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 were used as feature vectors for motion classification. The periodicity templates produced by the algorithm introduced here can provide both distinct shapes of regions of periodic motion, such as the wedge for the walking motion and the snow angle for the jumping jack, and accurate pixel-level description of a periodic action in the form of temporal harmonic energy ratio and motion fundamental frequencies.

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

5 Summary

We have presented a new method for detecting and segmenting multiple periodic motions in image sequences. The method consists of two main parts: 1) frame alignment to foveate on individual moving objects; 2) Fourier spectral harmonic peak detection and energy computation to identify regions of periodic motion. This method detects and segments a periodic motion simultaneously and is not computationally taxing. We illustrated the technique using real-world examples and demonstrated its robustness to noise.

A new periodicity measure in the form of temporal harmonic energy ratio is proposed. Compared to the periodicity measure used in other work, this is a more accurate quantitative measure of how much energy at each pixel location is contributed by the signal periodicity.

The periodicity templates contain information of the segmented region as well as the associated energy distribution and characteristic cyclic frequencies. The templates can be used as a means of periodic motion representation.

Acknowledgments

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

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. Int. Conf. on Patt. Rec.*, 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*, Austin, TX, 1994.

[4] R. Polana and R. C. Nelson. Detecting activities. In *Proceedings CVPR '93*, pages 2–7, New York City, NY, June 1993. Computer Vision and Pattern Recognition, IEEE Computer Society Press.

[5] A.F. Bobick and J.W. Davis. Real-time recognition of activity using temporal templates. In *Workshop in Applications of Computer Vision*, Sarasota, Florida, December 1996.

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

[7] 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.

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

[9] John Y. A. Wang. *Layered Image Representation: Identification of Coherent Components in Image Sequences.* PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, September 1996.

[10] J. Bergen *et al.* Hierarchial model-based motion estimation. In *Proc. Second Europ. Conf. Comput. Vision*, 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.