Real-Time Periodic Motion Detection, Analysis, and Applications

Ross Cutler and Larry Davis
University of Maryland, College Park
{rgc,lsd}@cs.umd.edu

Abstract

We describe a new technique to detect and analyze periodic motion as seen from both a static and moving camera. By tracking objects of interest, we compute an object's self-similarity as it evolves in time. For periodic motion, the self-similarity measure is also periodic, and we apply time-frequency analysis to detect and characterize the periodic motion. A real-time system has been implemented to track and classify objects using periodicity. Examples of object classification, person counting, and non-stationary periodicity are provided.

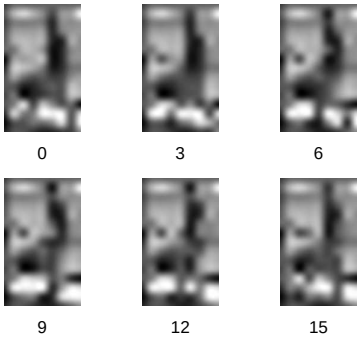

Figure 1. Low resolution image sequences of a periodic motion (a person walking on a treadmill). The effective resolution is 9x15 pixels.

1 Introduction

Object motions that repeat are common in both nature and the man-made environment in which we live. Examples include a person walking, a waving hand, a rotating wheel, and a flying bird. Knowing that an object's motion is periodic is a strong cue for object and activity recognition [4]. Periodic motion can also aid in tracking objects. Furthermore, the periodic motion of people can be used to recognize individuals [9].

We define the motion of a point $\vec{X}(t)$ at time t periodic if it repeats itself with a constant period p, i.e.:

$$\vec{X}(t) = \vec{X}(t+p) + \vec{T}(t), \tag{1}$$

where $\vec{T}(t)$ is a translation of the point. If p is not constant, then the motion is cyclic. In this work, we analyze locally (in time) periodic motion, which approximates many natural forms of cyclic motion.

1.1 Motivation

Our work is motivated by the ability of humans to recognize periodic movement at very low resolutions. For example, Figure 1 shows such a sequence. The effective resolution of this sequence is 9x15 pixels (it was created by resampling a 140x218 (8-bit, 30fps) image sequence to 9x15 and back to 140x218 using bicubic interpolation). In this sequence, note the similarity between frames 0 and 15. We will use image similarity to detect and analyze periodic motion.

1.2 Assumptions

In this work, we make the following assumptions: (1) the orientation and apparent size of the segmented objects don't change significantly during several periods (or do so periodically); (2) the frame rate is sufficiently fast to capture the periodic motion (at least double the highest frequency in the periodic motion).

2 Related Work

There has been recent interest in segmenting and analyzing periodic and cyclic motion. Existing methods can be categorized as those requiring point correspondences [16, 18]; those analyzing periodicities of pixels [10, 14]; those analyzing features of periodic motion [12, 3, 6]; and those analyzing the periodicities of object similarities [2, 16]. Related work has been done in analyzing the rigidity of moving objects [17, 11]. Below we review the most relevant

work.

Seitz and Dyer [16] compute a temporal correlation plot for repeating motions using different image comparison functions, d_A and d_I. The affine comparison function d_A allows for view-invariant analysis of image motion, but requires point correspondences (which are achieved by tracking reflectors on the analyzed objects). The image comparison function d_I computes the sum of absolute differences between images. However, the objects are not tracked, and thus must have non-translational periodic motion in order for periodic motion to be detected. The K-S test is utilized to classify periodic and non-periodic motion. When using d_I, this test fails to correctly classify most of our non-periodic data sequences (e.g., the sequence in Figure 11). In these cases, the correlation matrix is non-linear due to object viewpoint and lighting changes. The two samples being tested in the K-S test, the hypothesized period-trace and the correlation matrix, can have significantly different cumulative distribution functions; the motion is therefore incorrectly classified as periodic.

Liu and Picard [10] assume a static camera and use background subtraction to segment motion. Foreground objects are tracked, and their path is fit to a line using a Hough transform (all examples have motion parallel to the image plane). The power spectrum of the temporal histories of each pixel is then analyzed using Fourier analysis, and the harmonic energy cause by periodic motion is estimated. An implicit assumption in [10] is that the background is homogeneous (a sufficiently non-homogeneous background will swamp the harmonic energy). Our work differs from [10] and [14] in that we analyze the periodicities of the image similarities of large areas of an object, not just individual pixels aligned with an object. Because of this difference (and the fact that we use a smooth image similarity metric) our Fourier analysis is much simpler, since the signals we analyze do not have harmonics of the fundamental frequency. The harmonics in [10] and [14] are due to the large discontinuities in the signal of a single pixel; our self-similarity metric does not have such discontinuities.

Fujiyoshi and Lipton [3] segment moving objects from a static camera and extract the object boundaries. From the object boundary, a "star" skeleton is produced, which is then Fourier analyzed for periodic motion. This method requires accurate motion segmentation, which is not always possible (e.g., see Figure 8). Also, objects must be segmented individually; no partial occlusions are allowed (as shown in Figure 13). In addition, since only the boundary of the object is analyze for periodic change (and not the interior of the object), some periodic motions may not be detected (e.g., a textured rolling ball, or a person walking directly toward the camera).

This work extends our work in [2] by allowing a moving camera (rather than only a static one), and allowing for the detection and analysis of non-stationary periodicity (i.e., periodicity that changes in time) using time-frequency analysis. We also demonstrate that the object does not have be segmented from the background (if the background is sufficiently homogeneous). In addition, we explain the structure of the similarity matrix, and suggest how it can be exploited for object classification.

3 Method

The algorithm for periodicity detection and analysis consists of two parts. First, we segment the motion and track objects in the foreground. We then align each object along the temporal axis (using the object's tracking results) and compute the object's self-similarity as it evolves in time. For periodic motions, the self-similarity metric is periodic, and we apply time-frequency analysis to detect and characterize the periodicity.

3.1 Motion Segmentation and Tracking

Given an image sequence I_t from a moving camera, we segment regions of independent motion. The images I_t are first Gaussian filtered to reduce noise, resulting in I_t^*. The image I_t^* is then stabilized [5] with respect to image $I_{t-\tau}^*$, resulting in $\mathcal{S}_{t,t-\tau}$. The images $\mathcal{S}_{t,t-\tau}$ and I_t^* are differenced and thresholded to detect regions of motion, resulting in a binary motion image:

$$M_{t,-\tau} = \begin{cases} 1 \; if \;\; \left| I_t^* - \mathcal{S}_{t,t-\tau} \right| > T \\ 0 \; otherwise \end{cases} \tag{2}$$

where T is a threshold. In order to eliminate false motion at occlusion boundaries (and help filter spurious noise), the motion images $M_{t,\tau}$ and $M_{t,-\tau}$ are logically and'ed together, i.e., $M_t = M_{t,-\tau} \wedge M_{t,\tau}$. An example of M_t is shown in Figure 14. Note that for large values of τ, motion parallax will cause false motion in M_t. In our examples (for a moving camera), τ=300ms was used.

In many surveillance applications, images are acquired using a camera with automatic gain, shutter, and exposure. In these cases, normalizing the image mean before comparing images I_{t_1} and I_{t_2} will help minimize false motion due to a change in the gain, shutter, or exposure.

A morphological open operation is performed on M_t (yielding M_t^*), which reduces motion due to image noise. The connected components of M_t^* are computed, and small components are eliminated (further reducing image noise). The connected components which are spatially similar (in distance) are then merged, and the merged connected components are added to a list of objects O_t to be tracked. An object has the following attributes: area, centroid, bounding box, velocity, ID number, and age (in frames). Objects

in O_t and O_{t+k}, $k > 0$, are corresponded using spatial and temporal coherency.

It should be noted that the tracker is not required to be very accurate, as the self-similarity metric we use is robust and can handle tracking errors of several pixels (as measured in our examples).

3.2 Periodicity Detection and Analysis

The output of the motion segmentation and tracking algorithm is a set of foreground objects, each of which has a centroid and size. To detect periodicity for each object, we first align the segmented object (for each frame) using the object's centroid, and resize the objects (using a Mitchell filter [15]) so that they all have the same dimensions. The object O_t's self-similarity is then computed at times t_1 and t_2. While many image similarity metrics can be defined (e.g., normalized cross-correlation, [7], [1]), perhaps the simplest is absolute correlation:

$$S_{t_1,t_2} = \sum_{(x,y)\in B_{t_1}} |O_{t_1}(x,y) - O_{t_2}(x,y)|, \qquad (3)$$

where B_{t_1} is the bounding box of the object O_{t_1}. In order to account for tracking errors, the minimal S is found by translating over a small search radius r:

$$S'_{t_1,t_2} = \min_{|dx,dy|<r} \sum_{(x,y)\in B_{t_1}} \Big|O_{t_1}(x+dx, y+dy) - O_{t_2}(x,y)\Big|. \qquad (4)$$

For periodic motions, S' will also be periodic. For example, Figure 3 shows S' for all combinations of t_1 and t_2 for a walking sequence (the similarity values have been linearly scaled so that dark regions show more similarity). Note that a similarity matrix should be symmetric along the main diagonal; however, if substantial image scaling is required, this won't be the case. In addition, there will always be a dark line on the main diagonal (since an object is similar to itself at any given time), and periodic motions will have dark lines (or curves if the period is not constant) parallel to the diagonal.

To determine if an object exhibits periodicity, we analyze the spectral power of S' using Fourier analysis. While we could analyze the 2-D spectral power of S', for computational efficiency, we estimate the 1-D power spectrum of for a fixed t_1 and all values of t_2 (i.e. the columns of S'). In estimating the spectral power, the columns of S' are linearly detrended and a Hanning filter is applied. A more accurate spectrum is estimated by averaging the spectra of multiple t_1's [13] to get a final power estimate $P(f)$, where f is the frequency. Periodic motion will show up as peaks in this spectrum at the motion's fundamental frequencies. A peak at frequency f_p is significant if $P(f_p) > \mu_P + K\sigma_P$, where K is a threshold value (typically 3), m_P is the mean of P, and s_P is the standard deviation of P. Note that multiple peaks can be significant, as we will see in the examples.

3.3 Time-Frequency Analysis

For stationary periodicity (i.e., periodicity that doesn't change with time), the above analysis is sufficient. However, for non-stationary periodicity, Fourier analysis is not directly appropriate. Instead, we use time-frequency analysis and the Short-Time Fourier Transform (STFT) [13]:

$$F_x(t,v;h) = \int_{-\infty}^{\infty} x(u)h^*(u-t)e^{i2\pi vu}du, \qquad (5)$$

where $h^*(u-t)$ is a short-time analysis window, and $x(u)$ is the signal to analysis (S' in our case). The short-time analysis window effectively suppresses the signal $x(u)$ outside an neighborhood around the analysis time point $u = t$. Therefore, the STFT is a "local" spectrum of the signal $x(u)$ around t.

We use a Hanning windowing function as the short-time analysis window. The window length should be chosen to be long enough to achieve a good power spectrum estimate, but short enough to window a local change in the periodicity. In practice, a window length equal to several periods works well for typical human periodic motions (walking, running).

An example of non-stationary periodicity is given in Section 4.4.

3.4 Real-Time System

A real-time system has been implemented to track and classify objects using periodicity. The system uses a dual processor 400MHz Pentium II Xeon PC, and runs at 15Hz with 640x240 grayscale images captured from an airborne video camera. The system uses the real-time stabilization results from [5].

4 Examples

4.1 Person Walking on a Treadmill

The first example is of a periodic motion with no (little) translational motion, a person walking on a treadmill (Figure 2). This sequence was captured using a static JVC KY-F55B color camera at 640x480 @ 30fps, deinterlaced, and scaled to 160x120. Since the camera is static and there is no translational motion, background subtraction was used to segment the motion [2].

(a) (b) (c)

Figure 2. Person walking on a treadmill.

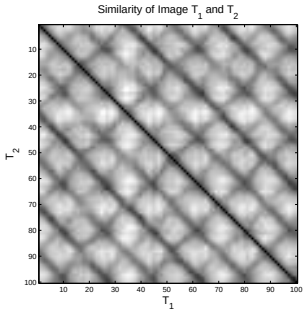

Figure 3. S' for a person walking on a treadmill. Dark regions correspond to images with greater similarity.

The similarity matrix S' for this sequence is shown in Figure 3; the first column of S' is shown in Figure 5. The darkest line is the main diagonal, since $S'(t,t) = 0$. The dark lines parallel to the main diagonal are formed since $S'(t, kp/2 + t) \approx 0$, where p is the period, and k is an integer. The dark lines perpendicular to the main diagonal are formed since $S'(t, kp/2 - t) \approx 0$, and is due to the symmetry of human walking. The cross diagonals causes a second peak in the power spectrum, double the fundamental frequency (Figure 6). It is interesting to note that at the intersections of these lines, these images are similar to either (a), (b), or (c) in Figure 2 (see Figure 4). That is, S' encodes the phase of the person walking, not just the period. This fact is exploited in the example in Section 4.3.

Note that not all periodic motion has an S' with cross-diagonal lines. For example, for the motion of a running dog with period p (captured at 30 fps), no two image frames within a cycle are similar (i.e., $S'(t_1, t_2) \gg 0, 0 \leq t_1 < t_2 \leq p$. This fact can be exploited for moving object classification.

4.2 Person / Vehicle Classification

A common task in an automated surveillance system is to classify moving objects. In this example, video from an airborne surveillance camera is used to track moving objects and classify them as humans or vehicles. While blob size

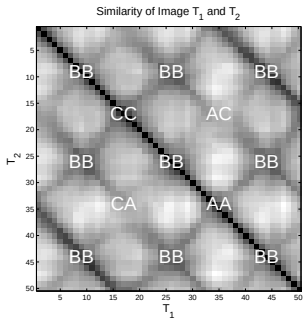

Figure 4. The intersections of S' are images similar to the poses given in Figure 2.

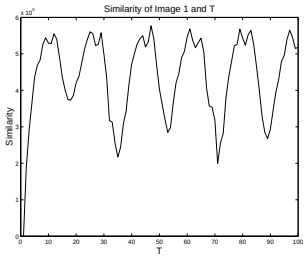

Figure 5. Similarity of image 1 with T for a walking person.

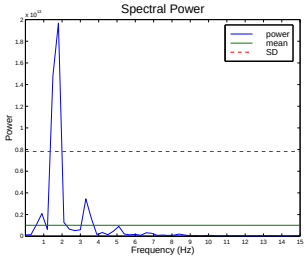

Figure 6. Average power spectrum of columns of S'.

Figure 7. Person running across a parking lot.

(a) (b) (c)

Figure 8. Zoomed images of the person in Figure 7, which correspond to the poses in Figure 2. The person is 7x12 pixels in size.

could be used for object classification, often the blobs are noisy, broken up, and don't accurately represent the object size. Further, blob size alone could incorrectly classify a group of people as a vehicle.

The video in this example was recorded from a Sony XC-999 camera (640x240 @ 30fps) at an altitude of 1500 feet. There is significant motion blur due to a slow shutter speed and fast camera motion. Additional noise is induced by the analog capture of the video from a duplicated SVHS tape. Figure 7 shows a person running across a parking lot.). The person is approximately 7x12 pixels in size (Figure 8). The similarity matrix in Figure 9 shows a clear periodic motion, which corresponds to the person running. Figure 10 shows that the person is running with a frequency of 1.3Hz; the second peak at 2.6Hz is due to the symmetry of human motion as described in Section 4.1. Figure 12 shows the similarity matrix for the vehicle in Figure 11, which has no periodicity. The spectral power for the vehicle is flat.

4.3 Counting People

Another common task in an automated surveillance system is to count the number of people entering and leaving an area. This task is difficult, since when people are close to each other, it is not always simple to distinguish the individuals. For example, Figure 13 is a frame from an airborne video sequence that shows three people running along a road; the result of the motion segmentation is shown in Figure 14. Simple blob counting will give an inaccurate es-

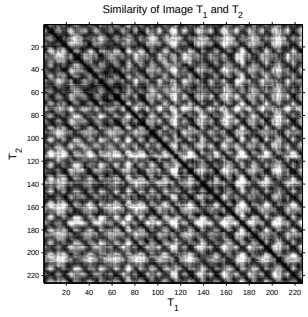

Figure 9. S' for the running person in Figure 7.

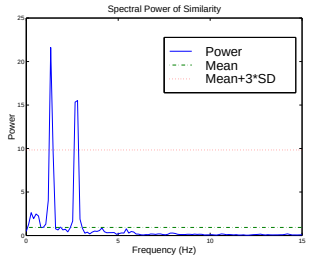

Figure 10. Spectral power the running person in Figure 7.

Figure 11. Vehicle driving across a parking lot.

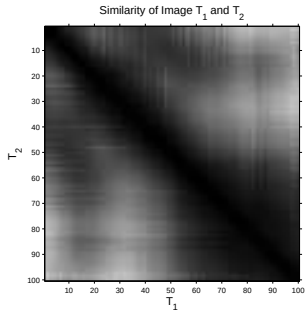

Figure 12. S' for the moving vehicle in Figure 11.

Figure 13. Three people running.

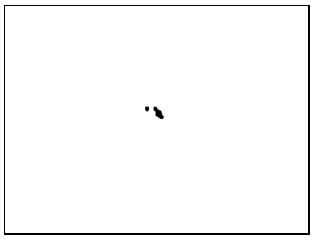

Figure 14. Segmented motion from Figure 13.

timate of the number of people. However, since we know the approximate location of the plane (via GPS), the approximate camera direction, and have an approximate site model (a ground plane), we can estimate the expected image size an "average" person should be. This size is used to window a region with motion for periodic detection. In this example, three non-overlapping windows were found to have periodic motion, each corresponding to a person. The similarity matrices and spectral powers are shown in Figure 15.

The similarity matrices in Figure 15 can also be used to extract the phase angle of the running person. In this example, the phase angles (which are encoded in the intersections of the main and cross-diagonal lines) are all significantly different from one another, giving further evidence that we have not over-counted the number of people.

4.4 Non-Stationary Periodicity

In this example, a person is walking, and roughly half way through the sequence, starts to run (see Figure 16). The similarity matrix (Figure 17) clearly shows this transition. Using a short-time analysis windowing Hanning function of length 3300ms (100 frames), the power is estimated in the walking and running stages (Figure 18).

5 Conclusions

We have described a new technique to detect and analyze periodic motion as seen from both a static and moving camera. By tracking objects of interest, we compute an object's

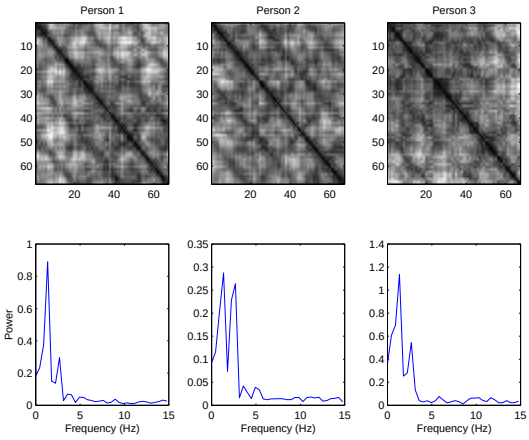

Figure 15. Similarity matrices and spectral power for the 3 people in Figure 13. Note that the frequency resolution is not as high as in Figure 10, since fewer frames are used to estimate the power.

Figure 16. Person walking, then running.

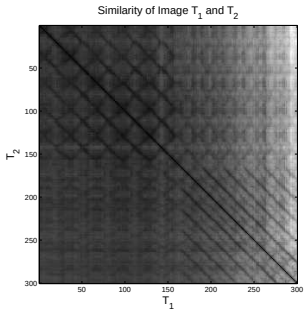

Figure 17. S' for the person walking/running in Figure 16.

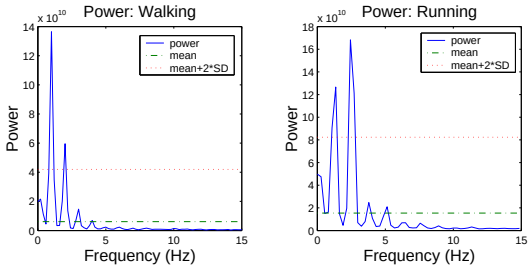

Figure 18. Spectral power for the person walking/running in Figure 16.

self-similarity as it evolves in time. For periodic motion, the self-similarity measure is also periodic, and we apply time-frequency analysis to detect and characterize the periodic motion.

Current work includes analyzing S' using texture analysis techniques (e.g., [8]) to robustly extract the parameters of the periodicity. We are also classifying different types of periodicity due to different types of motion symmetry (e.g., walking/running humans, running dogs, and flying birds all have distinguishing features in S' due to their different motion symmetries). Future work includes using alternative independent motion algorithms for moving camera video, which could make the analysis more robust for non-homogeneous backgrounds.

Periodicity can also be used to aid in tracking, particularly for occlusions (e.g., when two people walk past each other and one occludes the other), the periodicity (and image history) of the people can be used to determine which person is being occluded).

Acknowledgments: The airborne video was provided by the DARPA Airborne Video Surveillance project. This paper was written under the support of Contract DAAL-01-97-K-0102 (ARPA Order E653).

References

[1] D. H. Ballard and M. J. Swain. Color indexing. *Int. Journal of Computer Vision*, 7-1:11–32, 1991.

[2] R. Cutler and L. Davis. View-based detection and analysis of periodic motion. In *International Conference on Patern Recognition*, Brisbane, Australia, August 1998.

[3] H. Fujiyoshi and A. Lipton. Real-time human motion analysis by image skeletonization. In *IEEE Workshop on Applications of Computer Vision*, October 1998.

[4] N. H. Goddard. Human activity recognition. In M. Shah and R. Jain, editors, *Motion-Based Recognition*, pages 147–170. Kluwer Academic Publishers, 1997.

[5] M. Hansen, P. Anandan, K. Dana, G. van der Wal, and P. Burt. Real-time scene stabilization and mosaic construction. In *DARPA Image Understanding Workshop*, Monterrey, CA, Nov. 1994.

[6] B. Heisele and C. Wohler. Motion-based recognition of pedestrians. In *International Conference on Patern Recognition*, August 1998.

[7] D. Huttenlocher, G. A. Klanderman, and W. Rucklidge. Comparing images using the hausdroff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):805–863, 1993.

[8] H.-C. Lin, L.-L. Wang, and S.-N. Yang. Extracting periodicity of a regular texture based on autocorrelation functions. *Pattern Recognition Letters*, 18:433–443, 1997.

[9] J. Little and J. Boyd. Recognizing people by their gate: the shape of motion. *Videre*, 1(2), 1998.

[10] F. Liu and R. Picard. Finding periodicity in space and time. *International Conference on Computer Vision*, January 1998.

[11] D. McReynolds and D. Lowe. Rigidity checking of 3D point correspondences under perspective projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1174–1185, 1996.

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

[13] A. Oppenheim and R. Schafer. *Discrete-time signal processing*. Prentice-Hall, 1989.

[14] R. Polana and R. Nelson. Detection and recognition of periodic, non-rigid motion. *International Journal of Computer Vision*, 23(3):261–282, June/July 1997.

[15] D. Schumacher. General filtered image rescaling. In D. Kirk, editor, *Graphics Gems III*. Harcourt Brace Jovanovich, 1992.

[16] S. M. Seitz and C. R. Dyer. View-invariant analysis of cyclic motion. *International Journal of Computer Vision*, 25(3):1–23, 1997.

[17] A. Selinger and L. Wixson. Classifying moving objects as rigid or non-rigid without correspondences. In *DARPA Image Understanding Workshop*, pages 341–347, November 1998.

[18] P. Tsai, M. Shah, K. Keiter, and T. Kasparis. Cyclic motion detection. *Pattern Recognition*, 27(12):1591–1603, 1994.