HIGHER-ORDER SPECTRAL ANALYSIS OF HUMAN MOTION

A.N. Rajagopalan and R. Chellappa

Center for Automation Research
University of Maryland
College Park, MD 20742-3275

ABSTRACT

We describe a higher-order spectral analysis-based approach for detecting people by recognizing human motion such as walking or running. The periodic attribute of human motion lends itself to efficient spectral inspection. In the proposed method, the stride length is determined in every frame as the image sequence evolves. The bispectrum which is the Fourier transform of the triple correlation is a robust indicator of presence of periodicity. Triple correlation is robust as it is immune to any symmetrically distributed noise. The method is successfully tested on real video sequences.

1. INTRODUCTION

Detecting the presence of people in images is a very important and challenging problem because people are non-rigid objects and it is very difficult to describe them analytically. Typically, images of people present significant variabilities in color and texture patterns within the boundaries of the body. The increasing availability of video sensors and high performance video processing hardware opens up exciting possibilities for visual surveillance and multimedia applications.

Early approaches to human detection relied heavily on hand-crafted models [1, 2]. Representative works based on statistical learning principles may be found in [3] and [4]. Recently, there has been considerable interest in recognizing human motion by gait analysis [5] - [11]. These can be classified as motion-based recognition techniques. Repeated activity, such as human walking, tends to give rise to periodic or nearly periodic motion which is an important cue for recognition. By definition, a motion P is called periodic if it repeats with period p, i.e., $P(t+p) = P(t)$, for some constant $p > 0$ and all times t in a given time domain. The smallest such constant p is the period. When p is not constant, the motion is termed as cyclic.

Existing methods on motion-based recognition fall under one of the following categories: those requiring point correspondences; those analyzing periodicities of pixels; those analyzing features of periodic motion; and those analyzing the periodicities of object similarities. In [5], the motion of a walking person is analyzed by examining the cycles detected from spatio-temporal curvature functions of trajectories created by specific points on the object. Seitz and Dyer [6] use reflective markers to aid feature detection and tracking. An affine-invariant function is used to compare images over successive frames. A temporal correlation plot is computed and analyzed for periodic motion using the Kolmogrov-Smirnov test. In [7], Polana and Nelson propose a flow-based approach for motion detection. They extract reference curves, which are lines parallel to the trajectory of the motion centroid, and analyze the power spectra of these curves for periodicity. In the method proposed by Liu and Picard [8], the path of the tracked moving object is fit to a line using the Hough transform. The temporal history of each pixel is analyzed using the Fourier transform, along with a measure of periodicity in the form of harmonic energy ratios. In [9], moving targets are detected using motion information and a 'star' skeleton is produced from their boundaries. The motion of the skeleton segments is used to determine human activities such as walking or running. In [10], a clustering-based technique is proposed for detecting pedestrians in color images. In [11], an object's self-similarity is computed as it evolves in time and analyzed for periodicity.

We describe a robust higher-order spectral analysis-based approach for human motion detection. The method is based on a spectral examination of the motion of the boundaries of the object. Simple image operations are used to determine a bounding box, and hence a corresponding distance function, that captures the variations in the extremities of the moving object over time. Because the moment function of a periodic sequence is periodic, the bispectrum which is the Fourier transform of the triple-correlation of the distance function is examined for periodicity. Importantly, triple correlations are insensitive to any symmetrically distributed noise. Therefore, transformation to a higher-order domain reduces the effect of noise significantly. In a recent work [4], higher-order statistics have been very effectively used in a clustering framework for detecting people. However, unlike the method in [4], the approach presented here avoids training.

This work is partially supported by the ARL Federated Laboratory on Advanced Sensors Consortium under contract DAAL-01-96-2-0001.

0-7803-6297-7/00/$10.00 © 2000 IEEE

2. THEORY OF HIGHER-ORDER SPECTRA

Higher-order spectra are multi-dimensional Fourier transforms of higher-order statistics [12, 13]. We briefly discuss these concepts for the case of deterministic signals which includes both energy and power signals.

2.1. Energy Signals

Let $\{x(k)\}$ $k = 0, \pm 1, \pm 2, \ldots$ be a finite-energy deterministic signal. The n^{th}-order moments of $\{x(k)\}$, assuming they exist, are $(n-1)$-dimensional functions defined by

$$m_n^x(\tau_1, \ldots, \tau_{n-1}) \triangleq \sum_{k=-\infty}^{\infty} x(k)x(k+\tau_1)\ldots x(k+\tau_{n-1}) \,.$$

The n^{th}-order moment spectrum of $\{x(k)\}$ is defined as

$$M_n^x(\omega_1, \ldots, \omega_{n-1}) \triangleq \sum_{\tau_1=-\infty}^{\infty} \cdots \sum_{\tau_{n-1}=-\infty}^{\infty} m_n^x(\tau_1, \ldots, \tau_{n-1}) \cdot \exp\left(-j(\omega_1\tau_1 + \ldots + \omega_{n-1}\tau_{n-1})\right) \,.$$

Clearly, $M_n^x(\omega_1, \ldots, \omega_{n-1})$ is periodic with period 2π. The triple correlation, in particular, has an important advantage. It is insensitive to the presence of symmetrically distributed noise. Let the noise be additive, stationary and signal independent. The observation model is given by $y(t) = x(t) + n(t)$. It can be shown that for zero-mean noise,

$$< m_3^y(t_1, t_2) > = m_3^x(t_1, t_2) + < m_3^n(t_1, t_2) > \qquad (1)$$

where $m_3^y(\cdot)$, $m_3^x(\cdot)$ and $m_3^n(\cdot)$ denote triple-correlations of y, x and n, respectively, while $< \cdot >$ is the expectation operator. If the probability density function of the noise is symmetrical, i.e., $p(n) = p(-n)$, or at least not skewed, i.e., $\int n^3 p(n) dn = 0$, then the term $< m_3^n >$ is negligible which renders the triple correlation very effective in detecting a signal embedded in noise.

2.2. Periodic Signals

Consider a real power signal $\{x(k)\}$, $k = 0, \pm 1, \pm 2, \ldots$ The n^{th}-order moments of the power signal are defined by $r_n^x(\tau_1, \ldots, \tau_{n-1}) \triangleq < x(k)x(k+\tau_1)\ldots x(k+\tau_{n-1}) >$, where $< \cdot >$ corresponds to time-averaging operation. If the power signal is periodic with period N, that is, $\tilde{x}(k) = \tilde{x}(k+N)$, then

$$r_n^{\tilde{x}}(\tau_1, \ldots, \tau_{n-1}) = < \tilde{x}(k)\tilde{x}(k+\tau_1)\ldots\tilde{x}(k+\tau_{n-1}) > = \frac{1}{N}\sum_{k=0}^{N-1} \tilde{x}(k)\tilde{x}(k+\tau_1)\ldots\tilde{x}(k+\tau_{n-1}).$$

All the properties for moments of stationary random processes and energy signals also apply here for $r_n^x(\tau_1, \ldots, \tau_{n-1})$ and $r_n^{\tilde{x}}(\tau_1, \ldots, \tau_{n-1})$. Interestingly, an additional property of $r_n^{\tilde{x}}(\tau_1, \ldots, \tau_{n-1})$ that has *no counterpart* with moments of energy signals is that

$$r_n^{\tilde{x}}(\tau_1, \ldots, \tau_{n-1}) = r_n^{\tilde{x}}(\tau_1 + N, \ldots, \tau_{n-1} + N). \qquad (2)$$

That is, the n^{th}-order moment sequence of a periodic signal with period N is in turn periodic with the same period. The n^{th}-order moment spectrum of $\{\tilde{x}(k)\}$ is defined as

$$R_n^{\tilde{x}}(\lambda_1, \ldots, \lambda_{n-1}) = \sum_{\tau_1=0}^{N-1} \cdots \sum_{\tau_{n-1}=0}^{N-1} r_n^{\tilde{x}}(\tau_1, \ldots, \tau_{n-1}) \cdot \exp -j\frac{2\pi}{N}(\tau_1\lambda_1 + \ldots + \tau_{n-1}\lambda_{n-1})$$

where $R_n^{\tilde{x}}(\lambda_1, \ldots, \lambda_{n-1})$ is multidimensional periodic with period N.

3. BISPECTRUM-BASED DETECTION

In our approach to the detection of people in video, moving objects are detected by examining changes in the intensity values over image frames coupled with simple thresholding and morphological operations. Once a motion region is identified, the distance between the extremities of the object is approximately computed by fitting a bounding box to the motion region. If the moving object is a human, then this distance corresponds to the stride length. This step is repeated for every frame as the image sequence evolves with time. The triple correlation of the distance function is then computed and a periodicity detector examines the bispectrum to determine whether the motion actually corresponds to that of a human.

3.1. Segmentation

If the camera is assumed to be fixed, then the background scene is static. Hence, to find motion regions, the absolute difference of the current video frame from the previous one is computed to produce a difference image. Due to camera distortion, the background pixel values recovered by differencing may not be exactly the same from frame to frame. Also, one must allow for small changes in scene illumination with time. Hence, the difference image is thresholded at a suitable value to obtain a binary image. A simple thresholding scheme obviously has its own limitations. If the threshold is too low, camera noise and shadows will produce spurious objects; whereas, if the threshold is too high, some portions of the objects in the scene will fail to be separated from the background.

Moving sections are clustered using a connected component analysis. On the thresholded motion image, erosion is applied to the foreground pixels to eliminate small and spurious noise. This is followed by dilation to restore connectivity between components that are separated by a few pixels so as to yield the motion regions. In the case of a moving camera, the image frames must be first stabilized.

3.2. Periodicity Detection

The detected motion region could be the result of human motion or vehicular motion or 'nuisance' motion caused by small birds, vegetation, or the wind. The function of the detection module is to determine whether the foreground motion pixels actually correspond to those of a human in motion. The periodic attribute of human motion such as walking or running is a very useful and strong cue for recognition. It must also be noted that periodicity in these types of motion is primarily due to the legs. The rest of the body contributes minimally to the periodicity (except for the hands which might swing). When a person walks, the motion of the legs inevitably follows a regular pattern: as the person begins to walk, the legs initially move away from each other, then come close together, again move away from each other, only to come closer again. This motion cycle repeats itself systematically over time and is a stable and characteristic feature of human gait.

To determine whether the motion is periodic, we propose to examine the stride length which is a good indicator of the presence of periodicity in the motion. In Fig. 1, we show some representative output results of the segmentation algorithm. The motion regions correspond to the legs of a person. A bounding box is then fit so as to tightly enclose the detected motion region. This is achieved by continuously tightening the bounding box around the motion region until not more than 5% of the total motion pixels lie outside the bounding box. Experiments reveal that the toe is usually comprised of about 5% of the total motion pixels within the body. The dimension of the bounding box along the direction of motion is indicative of the stride length.

Fig. 1. Motion regions corresponding to the legs.

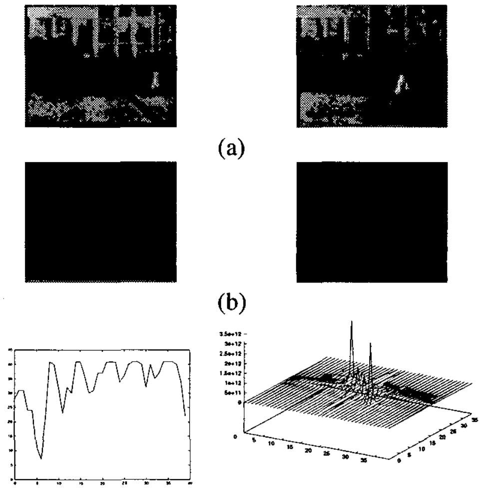
Let $x(n)$ denote the stride length in the n^{th} frame. Periodicity in the human gait is reflected in the signal $x(n)$ as it evolves with time. If $x(n)$ is periodic (or at least, nearly periodic), then according to the result in Section 2, the triple correlation of $x(n)$ would also be periodic. If the triple correlation is periodic, then the bispectrum would exhibit a relatively strong peak at the fundamental frequency (at which the person is walking). On the other hand, if $x(n)$ is not periodic, then the bispectrum will not have this characteristic. Thus, an examination of the bispectrum can be very useful in determining whether the moving object is a human or not.

Even after thresholding and morphological operations, the motion image may still contain spurious pixels. Because the stride length is determined using all the motion pixels within the detected region, the error due to spurious noise pixels is small. The bounding box has no bias towards positive or negative errors. Hence, the noise component of the distance computation module may be assumed to be approximately symmetric. The bispectrum is quite robust to symmetric noise and can tolerate errors up to several pixels.

4. EXPERIMENTAL RESULTS

In this section, we demonstrate the performance of the proposed bispectrum-based method for detection of human motion. As will be shown, the method can successfully differentiate between human motion (such as walking) and non-periodic activity (such as vehicular motion).

The first test video sequence consists of a man walking quickly across a road. Since the video was captured from a moving camera, the image sequences were first stabilized. Representative frames from the stabilized sequence are shown in Fig. 2(a). Background subtraction, followed by thresholding and morphology were then employed on the stabilized frames. From Fig. 2(b), we note that the resultant motion image contains the moving object (person, in this case) at the correct location in the image frames. A bounding box is fit to the detected motion region in each frame and the resultant stride length is plotted in Fig. 2(c). The periodic nature of the stride length is clearly evident from the figure. However, it should be noted that the curve is not regular. This

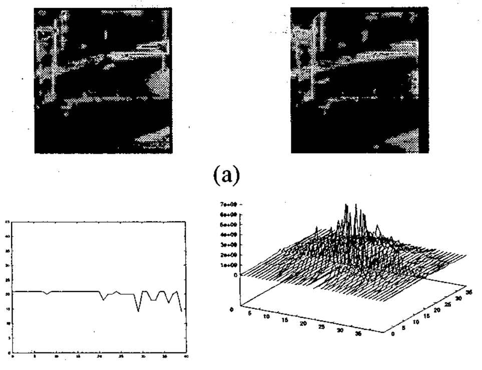


(a)

(b)

(c) (d)

Fig. 2. (a) Stabilized image frames from a walking sequence. (b) Output of the motion detection module. (c) Stride length versus frame number. (d) Bispectral plot of the function in (c).



(a)

(b) (c)

Fig. 3. (a) Stabilized images of a car sequence. (b,c) Plots corresponding to the distance function and its bispectrum.

is because human motion is seldom perfectly periodic. Also, some errors are inevitable in the computation of the bounding box. Because the person was walking quickly, there are about 6 cycles of motion over 42 frames. The triple-correlation of the distance curve was next computed. It is very interesting to note that despite the presence of noise in the distance curve, the bispectrum (Fig. 2(d)) exhibits a strong peak at exactly $\lambda_1 = \lambda_2 = 6$, which is the fundamental frequency.

In the next experiment, a test sequence with non-periodic activity was chosen. The sequence consisted of a car in motion captured from a non-stationary platform. A few of the motion compensated frames are shown in Fig. 3(a). The moving car was located in each image frame by using frame differencing and morphological operations. The distance function was computed for each frame and is plotted in Fig. 3(b). Clearly, the curve does not reveal any kind of systematic repetitive activity. The bispectrum was computed from the triple-correlation of the distance function and is plotted in Fig. 3(c). As expected, the bispectrum does not exhibit any relative strong peaks, unlike the case of human motion. Thus, the proposed method can effectively discriminate between motion due to people and vehicles.

5. CONCLUSIONS

We have described a higher-order spectra-based method for detecting people by recognizing human motion such as walking or running. In this method, the distance between the extremities of the moving object is determined in each frame. Because of the specific characteristics of human motion, the stride length tends to exhibit periodicity. We use the important result that for a periodic sequence, the higher-order moment sequence is also periodic. An examination of the bispectrum reveals whether the moving object is a human in motion. The system was tested on real image sequences and was found to successfully differentiate between people and vehicular motion. For a more formal examination of the bispectrum, we are currently working on bispectrum-based hypothesis test for periodicity within a rigorous statistical framework.

Acknowledgments

The test image sequence of the walking-man was kindly provided by Dr. Bernd Heisele, Center for Biological and Computational Learning, MIT.

6. REFERENCES

[1] M.K. Leung and Y.H. Yang, "Human body motion segmentation in a complex scene, *Pattern Recognition*, vol. 20, no. 1, pp. 55-64, 1987.

[2] K. Rohr, "Incremental recognition of pedestrians from image sequences", *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition* (New York, USA), 1993, pp. 8-13.

[3] C.P. Papageorgiou, M. Oren and T. Poggio, "A trainable system for people detection", *Proc. DARPA Image Understanding Workshop*, (New Orleans, LA), May 1997, pp. 207-214.

[4] A.N. Rajagopalan, Philippe Burlina and R. Chellappa, "Detection of people in images", *Proc. Intl. Joint Conf. on Neural Networks* (Washington, D.C), July 1999.

[5] P. Tsai, M. Shah, K. Keiter and T. Kasparis, "Cyclic motion detection for motion based recognition", *Pattern Recognition*, vol. 27, pp. 1591-1603, 1994.

[6] S.M. Seitz and C.R. Dyer, "View-invariant analysis of cyclic motion", *International Journal of Computer Vision*, vol. 25, pp. 231-251, 1997.

[7] R. Polana and R.C. Nelson, "Detection and recognition of periodic, nonrigid motion", *International Journal of Computer Vision*, vol. 23, pp. 261-282, 1997.

[8] F. Liu and R.W. Picard, "Finding periodicity in space and time", *Proc. IEEE Intl. Conf. on Computer Vision*, 1998, pp. 376-383.

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

[10] B. Heisele and C. Wohler, "Motion-based recognition of pedestrians", *Proc. IEEE Intl. Conf. on Pattern Recognition*, 1998.

[11] R. Cutler and L. Davis, "Real-time periodic motion detection, analysis, and applications", *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition* (Fort Collins,Colorado), 1999, pp. 326-332.

[12] C.L. Nikias and J.M. Mendel, "Signal processing with higher-order spectra", *IEEE Signal Processing Magazine*, pp. 10-37, 1993.

[13] A.W. Lohm and B. Wirnitzer, "Triple Correlations", *Proceedings of the IEEE*, vol. 72, pp. 889-901, 1984.