Segmentation of Periodically Moving Objects

Ousman Azy and Narendra Ahuja
Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign, USA
{oazy2,n-ahuja}@uiuc.edu

Abstract

We present a new approach for the identification and segmentation of objects undergoing periodic motion. Our method uses a combination of maximum likelihood estimation of the period, and segments moving objects using correlation of image segments over an estimated period of interest. Correlation provides the best locations of the moving objects in each frame. Segmentation tree provides the image segments at multiple resolutions. We ensure that children regions and their parent regions have the same period estimates. We show results of testing our method on real videos.

1. Introduction

Suppose we are given a video sequence acquired from a static camera and containing objects moving nearly periodically in the image, with unknown periods. We wish to segment the moving objects and estimate their periods.

Existing methods for analyzing periodic motions focus on the estimation of periods. Period detection based on autocorrelation is performed on trajectories given by markers in [7]. Instead of reflective markers, feature points and their local image properties are used in [6]. Cross correlating every pair of frame in the spatial domain as presented in [4] creates a periodic pattern that can be used for recognition. In the work on segmentation of periodic motion by Briassouli et al.[3], the motion field is inferred using Fourier phase analysis and harmonic analysis. A block-wise correlation between T-spaced frames is performed for segmentation. Correlation is adversely affected by any appearance changes due to illumination changes and occlusion.

In this paper, we use a locally optimal period detector for segments present in a multiscale segmentation tree proposed in [1], while enforcing spatial-temporal consistency.

2 Algorithm

2.1 Detection and Estimation of the period

At each pixel, we perform hypothesis testing to estimate the period of its temporal variation. We treat the image signal as undergoing "epoch folding" introduced in [5], i.e., "folding" the 1D signal x along its assumed period T. The folded signal y is defined on the interval $0 < i < T$ as:

$$y_i = \frac{1}{\frac{N}{T}} \sum_{k=1}^{\lceil \frac{N}{T} \rceil} (x_{i+(k-1)T} - \bar{x})$$

If the signal is not T-periodic then folding will yield a signal approaching uniform noise. However if the signal is T-periodic, the shape of overlaying parts of the signal is identical. To distinguish between those two situations, Pearson's chi square test statistic used by determining whether y after folding is uniform or not:

$$\chi^2 = \sum_{k=1}^{T} \frac{(y_k - \bar{y})^2}{\bar{y}}$$

where χ^2 follows a χ^2 statistic with degrees of freedom being $T-1$. The likelihood that the pixel period is T is proportional to $\chi^2(T)$. this is the best estimator of the period of a periodic signal in the maximum likelihood sense, as proven in [8].

Next, we extend the above, pixel-based period estimation to regions of pixels. For simplicity, we assume that the pixels of a segment move independently:

$$p(T_{interior,border}) = \prod_{pixels \in interior,border} p(T_{pixel})$$

where $p(T_{interior})$ and $p(T_{border})$ are, respectively, the probability densities of the period of the pixels inside the children subregions of a given region, and the remaining pixels, contained in the outer shell that remains after the children subregions are removed from the region. Given no priors, the distribution of the period is

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

uniform. Thus, the MAP rule becomes the maximum likelihood test. This assumption is increasingly valid for smaller regions. Thus we obtain the following optimal estimate of the period for each pixel :

$$\hat{T}_{interior,border} = argmax(p(T_{interior,border}))$$

$$\hat{T}_{region} = max(p(T_{interior}), p(T_{border})))$$

The purpose of distinguishing between the period of the pixels in the border of a region and the period of the pixels of the interior is to handle the common case of a region with uniform color where the motion manifested by a variation of the intensity is more apparent in the border with the background than in the interior.

To segment all the pixels moving with the same period, i.e., belonging to the same object, we use Fourier analysis. Using the fact that the Fourier Transform of a periodic signal is a discrete spectrum of samples taken at time instants that are multiples of $\frac{1}{T}$, we extract the periodic components of a periodic signal by filtering the corresponding frequencies. To increase accuracy, we zero pad the sequence such that multiple of $\frac{1}{T}$ matches a value multiple of $\frac{1}{numberofframes}$ samples. This enhances the frequency resolution at the expense of computational complexity.

Static background is subtracted by median filtering. Given a period of interest T, the corresponding, segments in the frame at time t and $t+T$ are correlated. A high correlation means a high confidence that the moving segment has period T.

2.2 Segmentation tree for spatial and temporal consistency

The previous detectors indicate the positions of the objects in motion. For their segmentation, we use the segmentation tree, which consists of segments occurring in the image at all color resolutions, as well as their mutual containment relationships captured by parent-child links in the tree [2]. Then within each region in the segmentation tree, the density of the detector response is computed:

$$density(region) = \frac{\iint_{region} detector(x,y)\, dx\, dy}{\iint_{region} dx\, dy}$$

where the periodicity detector operates on pixels.

We apply the period estimation process recursively, starting with the leaves, until we reach the highest level giving an acceptable estimate.

Given a frame, the three detectors namely those obtained by epoch folding, Fourier analysis, and correlation outputs are normalized and equally weighted. A k-means clustering with $k = 2$ for background/periodic foreground segmentation is used to merge the data.

Segments from the set of subtrees detected as moving periodically are tracked between two adjacent frames using bipartite graph matching. The similarity matrix is based on the Euclidian distance between the regions-node properties. An optimal estimate of the period is obtained using the Hungarian algorithm. From the displacement between frame t and frame $t+1$, we perform a partial cumulative likelihood test over time t and time $t+1$, over the two corresponding covered matched regions.

$$p(T_{interior,border_{[t,t+1]}}) = \prod_{pixels \in interior,border_{[t,t+1]}} p(T_{pixel})$$

we obtain a new estimate of the period:

$$\hat{T}_{interior,border_{[t,t+1]}} = argmax(p(T_{interior,border_{[t,t+1]}}))$$

$$\hat{T}_{region_{[t,t+1]}} = max(p(T_{interior_{[t,t+1]}}), p(T_{border_{[t,t+1]}}))$$

3 Experiments

We tested our algorithm on real sequences containing periodic motions: gym exercise sequences of resolution 160-by-120 pixels taken from a static camera. The results presented in Table 1 show that we obtain reasonable estimates of the periods of the motions present in the video.

Table 1. Estimation of the period

sequence	ground truth period	detected period
leg	32.33	29
diagonal	20	20
traction	23	21

In Fig. 1, the leg/weight/shoe video is correctly segmented throughout the video. Various periodic motion patterns are detected. Even segments with uniform intensity like the shadow on the floor can be detected. The quality of the period estimate is limited by the level (degree of detail) of segmentation used: the knee may be wrongly detected because of the low contrast with the background prevents the segment from being separated from its surround. Compared to a simple background subtraction method, our approach recovers the whole moving object irrespective of the amplitude of the motion. In the "traction" sequence, in which a person is pulling a wire attached to the left arm of the machine back and forth, presents many challenges. Many different periodically moving objects, such as the fans appear as blurred patches in the ceiling (moving with a period of 11 frames) in the back of the scene, and the wire/joint/person/weight (period 20 frames) are detected. The detection is robust to self occlusion of the arm on the body, but can be limited by the size of the

supporting segments: the joint of the wire and machine is detected but not the wire which is too small. In the "diagonal" sequence, in which a person is pulling the wire in diagonally, the main moving part, the upper body, is correctly segmented across the sequence. Our method performs perceptually better than the results in [3] as shown in the three last rows of Fig. 1. A segmentation error analysis is presented in Table 2. A quantitative measure of the error in the estimated segmentation is the ratio of the areas of the XOR between our mask and the ground truth, and the area of the union of the two computed across one period. Our error is around 25% of accuracy according to this measure while the human judgment error is around 5%. The segmentation error may also be evaluated as a function of the size of the region (related to detail captured at the associated segmentation tree level). Given a frame, each region of the segmentation tree is classified by its labeled with its size. For a random frame in 'diagonal", Fig. 2 confirms our intuition: most of the error is contributed by the smallest regions. Probability of miss and false alarm are treated equally in computing the segmentation error. A more detailed analysis is provided in Fig. 3, where the segmentation error is shown as a function of the frames number, for 20 frames. In the beginning, the error can be large but it quickly drops to a more reasonable level thanks to the temporal propagation mechanism.

Table 2. Average segmentation error in %

sequence	P_{error}	$P_{missing}$	$P_{falsealarm}$
leg	24.22	3.41	20.81
diagonal	28.84	7.6	21.23
diagonal (in [3])	57.34	28.29	29.04
traction	20.71	8.17	12.32

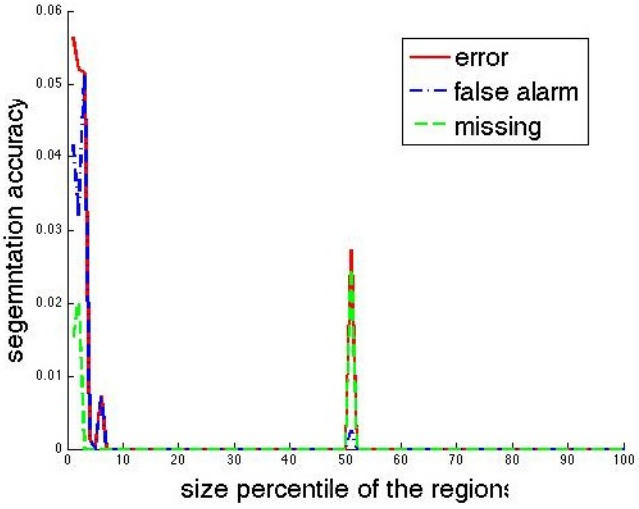

Figure 2. Segmentation error per segments size

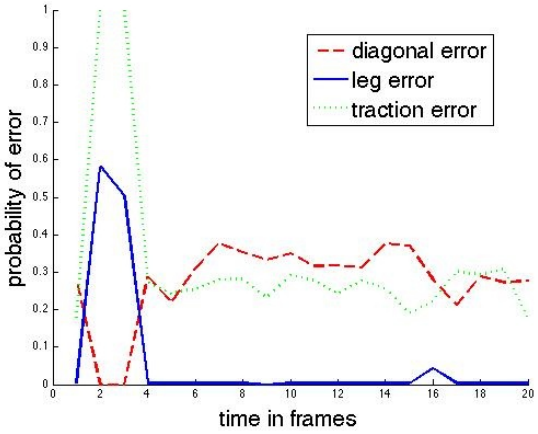

Figure 3. Time evolution of the segmentation error

4 Conclusion

Future work will focus on coarse-to-fine period estimation by traversing the segmentation tree top-down, through analysis of increasingly smaller regions. Our method could also be extended to periodic motions superposed with other motions such as translation and camera motions.

5 Acknowledgement

The support of the National Science Foundation under grant NSF IBN 04-22073 is gratefully acknowledged.

References

[1] N. Ahuja. A transform for multiscale image segmentation by in integrated edge and region detection. *PAMI*, 18(12):1–100, December 1996.

[2] H. Arora and N. Ahuja. Analysis of ramp discontinuity model for multiscale image segmentation. *CVPR*, 99(7):1–100, January 2006.

[3] A. Briassouli and N. Ahuja. Extraction and analysis of multiple periodic motions in video sequences. *PAMI*, 7(7):1244–1261, July 2007.

[4] R. Cutler and R. Eagleson. Robust real-time periodic motion detection, analysis, and applications. *PAMI*, 22(7):781–796, August 2000.

[5] D. Leahy, R. Elsner, and M. Weissskopf. On searches for pulsed emission: The rayleigh test compared to epoch folding. *ApJ*, 272(7):256–258, January 1983.

[6] R. Polana and R. Nelson. Detection and recognition of periodic, nonrigid motion. *IJCV*, 23(3):261–282, July 1996.

[7] P. Tsai, M. Shah, K. Keiter, and T. Kasparis. View-invariant analysis of cyclic motion. *IJCV*, 25(12):1–23, January 1997.

[8] J. Wise and T. Parks. Maximum likelihood pitch estimation. *ASSP*, 24(7):418–423, October 1999.

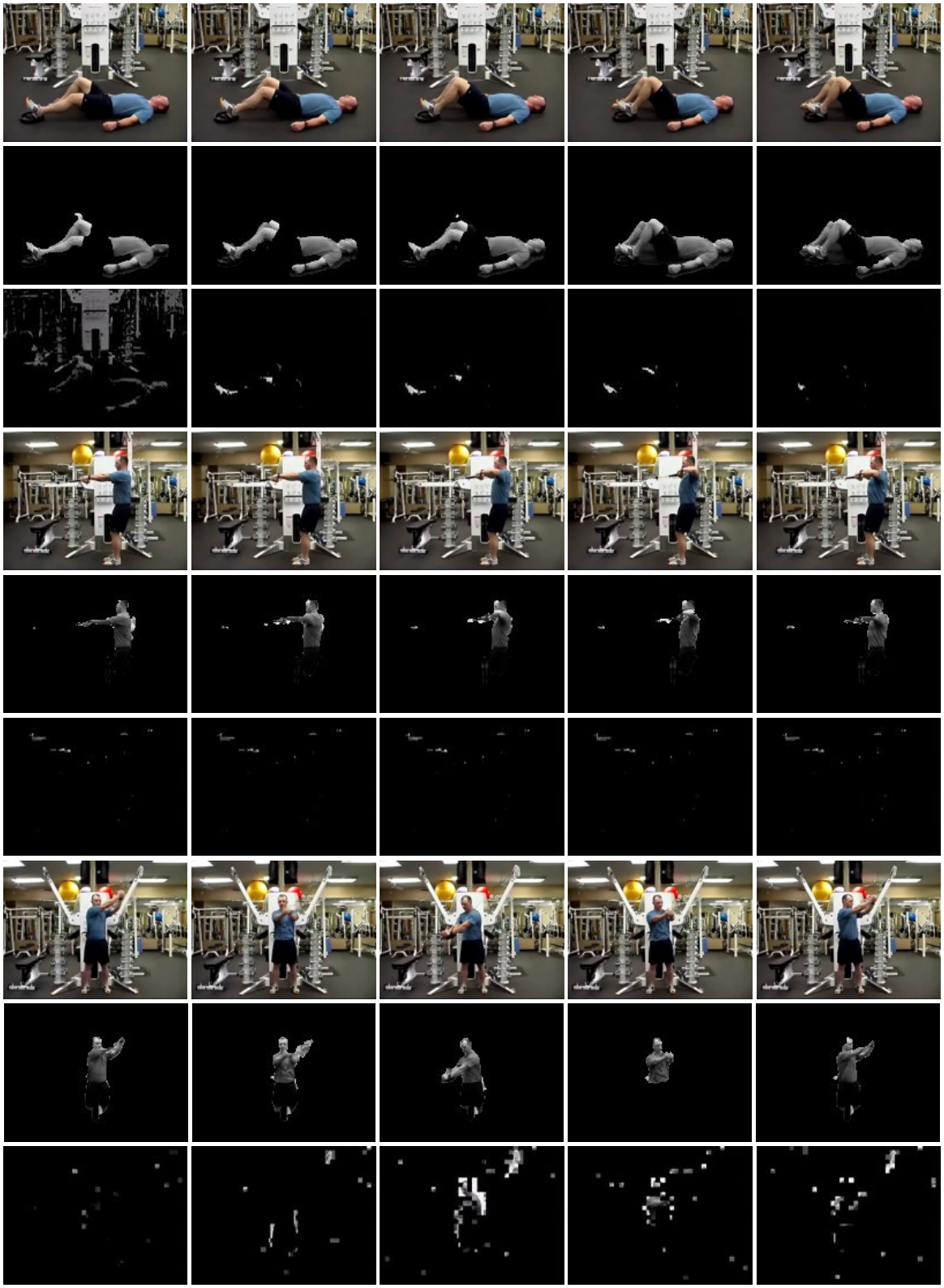

Figure 1. Final periodic motion segmentation for "leg" sequence (row 2) compared to that obtained using background subtraction (row 3). Segmentations of multiple periodic motions are shown for "traction" sequence: person moving at $T = 20$ (row 5), fans at $T = 10$ (row 6). For 'diagonal" (row 8) our segmentation is compared to that of [3] (row 9). Every 5th frame of the sequences is shown.