Periodic Motion Detection and Estimation via Space-Time Sampling

Ashwin Thangali and Stan Sclaroff [*]
Computer Science Department
Boston University, Boston, MA 02215
email: {tvashwin,sclaroff}@cs.bu.edu

Abstract

A novel technique to detect and localize periodic movements in video is presented. The distinctive feature of the technique is that it requires neither feature tracking nor object segmentation. Intensity patterns along linear sample paths in space-time are used in estimation of period of object motion in a given sequence of frames. Sample paths are obtained by connecting (in space-time) sample points from regions of high motion magnitude in the first and last frames. Oscillations in intensity values are induced at time instants when an object intersects the sample path. The locations of peaks in intensity are determined by parameters of both cyclic object motion and orientation of the sample path with respect to object motion. The information about peaks is used in a least squares framework to obtain an initial estimate of these parameters. The estimate is further refined using the full intensity profile. The best estimate for the period of cyclic object motion is obtained by looking for consensus among estimates from many sample paths. The proposed technique is evaluated with synthetic videos where ground-truth is known, and with American Sign Language videos where the goal is to detect periodic hand motions.

1. Introduction

The primary motivation of this paper is to detect and localize periodic movements in sign language video. Periodic motions are used for a multitude of linguistic purposes in signed languages, both in manual signing and in head movements, such as nods and shakes. With respect to manual signing, many signs are inherently periodic. In addition, periodicity (specifically, partial – and sometimes cyclic – reduplication of the sign, one or more times, in conjunction with other movement of the hands) is frequently used to express grammatical inflections, such as aspect (e.g., continuative, frequentative) or plurality. Detecting periodic motions also has important applications in surveillance and human gait analysis. The distinctive features of our technique are that we need neither object segmentation nor feature tracking, which are major drawbacks of previous techniques. We demonstrate good performance with low reso-

lution 160×120, 15 fps video, an example of video used in our experiments is given in Fig. 4.

The premise of our approach is that intensity values along a linear space-time sample path show a characteristic pattern if an object with periodic motion intersects this path. As the object cuts across a sample path, it induces oscillations in intensity values. A bright object (e.g. a hand) moving against a dark background induces intensity peaks at intersection points.

Note that the pattern of peaks induced in a randomly chosen sample path are not necessarily periodic. To illustrate this, consider the example of an oscillating sphere. The sphere traces a sine wave in space-time as shown in Fig. 1(b). The sample paths at different orientations are displayed by dashed lines and the corresponding intensity patterns along these paths are shown in (c)-(f). The intensity profiles are clearly *aperiodic* and hence a frequency analysis of these profiles to determine the dominant frequency would not be useful. *The induced intensity pattern is a function of parameters of cyclic motion (period and phase) as well as the orientation of the sample path (slope and offset) with respect to the axis of cyclic motion.*

In this paper, we present a least squares minimization technique to recover these parameters for each intensity profile. In the experiments section Sec. 4.1, we demonstrate that our technique is applicable to real videos and show results for different kinds of periodic motion commonly observed in sign language. The steps in our approach are as follows,

1. Choose a window of frames in the video. In our implementation, we use successive fixed length windows and analyze each window independently for evidence of periodicity.
2. Extract motion information from the first and last frames (boundary frames) of the window.
3. Choose sample points independently from the boundary frames using motion magnitude as a density function.
4. Obtain sample paths in space-time connecting pairs of sample points from the boundary frames and extract intensity values along these paths.
5. Extract intensity peaks in each path. Use the peak locations to obtain an estimate of parameters of object motion and sample path orientation with respect to object motion.
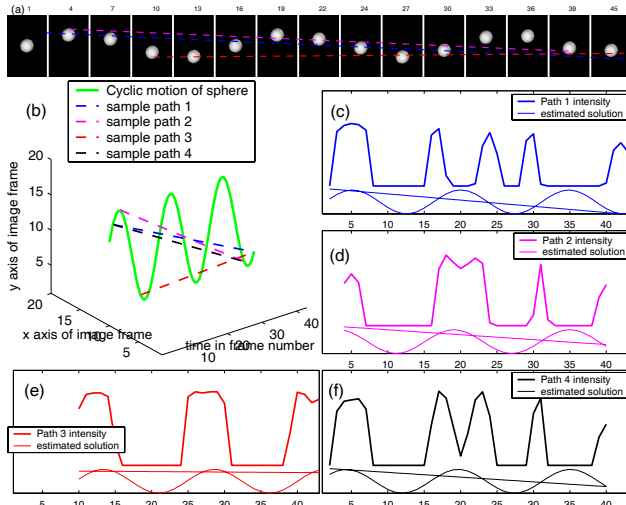
Figure 1: Synthetic example to illustrate intensity patterns for sampling from periodic motion. (a) shows frames from video of a sphere reciprocating vertically. (b) shows the motion of the sphere and orientation of sample paths in *space-time*. (c)-(f) show the intensity values sampled by the sample paths. Note that the intensity profiles are not necessarily periodic and a frequency analysis to extract the dominant frequency would be inaccurate. The solution for parameters of cyclic motion and sample path orientation estimated by the proposed technique are shown below each intensity profile, the estimated parameters match the true value.

6. Refine the parameter estimates using all intensity values along the sample path.
7. Use consensus in period estimates from many paths to obtain the best period estimate and a set of good candidate paths.
8. Re-sample additional sample paths oriented close to the candidate paths in space-time and re-run the estimation process. Use the most frequent period as an estimate of period for the current frame window.

1.1. Assumptions in our approach

In the following discussion we use the centroid of an object to characterize its motion. The periodic motion of an object (e.g., hands in sign language video) can be approximated by a periodic local motion component superimposed on a aperiodic (global) motion component.

1. **Aperiodic or global motion:** The projected global motion of objects on the image plane is approximately linear, i.e. the objects move along nearly straight lines in the image sequence.
2. **Periodic motion:** Objects show periodic movement in some fixed direction in the image plane, i.e., the periodic local motion component when projected onto the image plane is well approximated by reciprocating sinusoidal motion in some direction. Note that the direction of periodic motion is independent of the global

motion. We have observed that most types of cyclic human motion do project to reciprocating motions that are well approximated by sinusoidal motion in the image plane. Pendulum motions (motion of limbs of pedestrians or animals) are approximated quite accurately as reciprocating sinusoidal motions when the radius is large compared to the amplitude of motion. Circular or elliptical motions in a plane perpendicular to the image plane also project to (approximately) reciprocating sinusoidal motion on the image plane.

3. **Multiple objects:** We explicitly model the intersections of a sinusoidally reciprocating object with linear sample paths. We are hence unable to handle cases where multiple objects intersect a sample path. If the objects do not occlude each other, our technique can still be used since we expect that among a large number of sample paths, a significant number would only sample from one of the objects.
4. **Background:** A periodic texture on the background would influence period estimates obtained for paths that sample from the background. If the background is static, a background subtraction technique can be applied. To handle dynamic backgrounds, color information can be used to extract foreground regions [9]. In our experiments, hands have good contrast against the background and the background is smooth. The intensity fluctuations in the background hence did not significantly affect the estimation.

1.2. Related work

Techniques proposed in the past to detect periodicity assume a model (typically linear) for global motion and use either tracking or space-time analysis to estimate the parameters of global motion. A class of techniques use estimated global motion to align successive frames so as to reduce the periodicity detection problem to a 1D frequency analysis problem. Among these techniques, Niyogi and Adelson [1] use XT space-time slices (horizontal slices through space-time at different heights) to detect 2D periodic motions in the horizontal plane, e.g. legs of walking people.

This work was further extended by Liu and Picard [2] who use the Hough transform on XT slices to estimate linear translational motion and align successive frames to compensate for the translation of objects. They require that some parts of the object *do not* exhibit periodic motion for the alignment to be successful (clearly the Hough transform would perform well in a slice taken at the head and would fail for a slice at the legs). They then extract peaks in the frequency spectra of 1D samples in time for each pixel location. Since the frames are aligned, the sample intensities are periodic and hence a frequency domain technique is directly applicable. They show results on walking motion. It is not

clear if the Hough transform alignment would be applicable for complicated 3D periodic motions of hands.

Techniques based on correlation between pairs of aligned frames separated in time have been proposed by Cutler, et.al. [4] and Seitz, et.al. [3]. These techniques are based on the assumption that pairs of aligned frames separated by the right period correlate well. In our low quality sign language video, the hands are mostly featureless; Hence, we would not expect to see much difference in correlation values of frames separated by correct and incorrect period values.

A second class of techniques relies on tracking features on moving objects. Ormoneit, et.al. [5] use a correlation technique on estimated articulated body parameters to detect periodicity. Davis, et.al. [6] track features and analyze their trajectories for periodic motion in the image plane. Xion, et.al. [7] track hand location and use a wavelet approach to find dominant frequencies from hand trajectories. These techniques are highly susceptible to tracking errors and are inappropriate in our setting since robust tracking in low resolution video is still very much an open research problem.

2. Sampling paths in space-time

Task: Given a video segment, extract linear sample paths in space-time that sample from moving objects.

We require that each object be adequately sampled in space-time, i.e that there are a good number of sample paths for each object. It is likely that many paths sample from multiple objects; we can detect and reject such paths in later steps. The input window of frames should be small to ensure that the aperiodic component of object motion is linear but large enough to capture adequate number of periods (at least two periods).

The steps to obtain sample paths are as follows. We first extract motion information for the first and last frames (boundary frames) using a MPEG-type block based correlation technique as shown in Fig. 2(a). We use the motion magnitude as a density function to draw samples from pixels with high motion for both boundary frames. Sampling points from a density function is popularly known in literature as *importance sampling*. We add an additional constraint to ensure that each moving object is adequately sampled: the set of sample points should be chosen so as to maximize pairwise distances. To achieve this, we first bin pixels in the image based on their magnitude of motion (we only consider pixels with motion magnitude exceeding a threshold) so that each bin has roughly the same number of pixels. We then sample equal number of pixels from each bin by adding new points farthest from the current set of sample points. In our experiments, we found that 10 sample points are usually sufficient for two moving objects, e.g.

a pair of hands. We then extract intensity values along *linear* space-time sample paths connecting all pairs of samples between the boundary frames as shown in Fig. 2(a). Note that we use the linear global motion assumption in Sec. 1.1 to draw samples along straight lines in space-time. Also note that no tracking information is needed to sample paths.

3. Estimating period of cyclic motion

Task: To determine the period of cyclic motion given intensity values along a linear space-time sample path.

3.1. Model for samples from periodic motion

The sampling scheme discussed in the Sec. 2 compensates for global trend in object motion. The residual motion component corresponds to local perturbations that may either be periodic or aperiodic; we wish to distinguish between them and also estimate the period in the cyclic case. In Sec. 1.1, we have assumed that the projected periodic motion in the image plane is a sinusoidal reciprocating motion in some *unknown* direction d in the image plane. Consider a 1-d reference frame with x-axis aligned with d, we can approximate the periodic motion of the object in this reference as $x_o(t) = sin(\omega t + \phi)$.

A sample path linear in time is given by, $x_s(t) = at + b$. We assume that the sample paths are in the same plane as that of object motion. We have found that this is reasonable in practice. Consider the case of a bright object whose size is small compared to the amplitude of motion. We would see peaks in the intensity values at times t_i where the sample path intersects the object path and are given by $sin(\omega t_i + \phi) = at_i + b$. An example intensity profile from a sign language video in Fig. 2(b) clearly shows such peaks. In the case of a dark object against a bright background we can achieve the same by inverting the intensity values.

3.2. Overview of the estimation process

We run a peak detector on the intensity values and use the peak locations in a least squares formulation to estimate the unknown parameter values. This is discussed in Sec. 3.3 and the solution obtained illustrated in Fig. 2(c). If the size of the object is comparable to the amplitude of motion, the peaks would be blurred considerably and some peaks would be lost. In practice, we would have considerable noise since real objects (e.g. hands) are seldom smooth. Peak detection would hence not be robust. In Sec. 3.4, we formulate a least squares solution utilizing all the sample values instead of only the peak locations. A solution of the continuous formulation is shown in Fig. 2(e). Given the parameter estimates from many paths, Sec. 3.5 describes a consensus and re-sampling scheme to obtain the best period estimate.
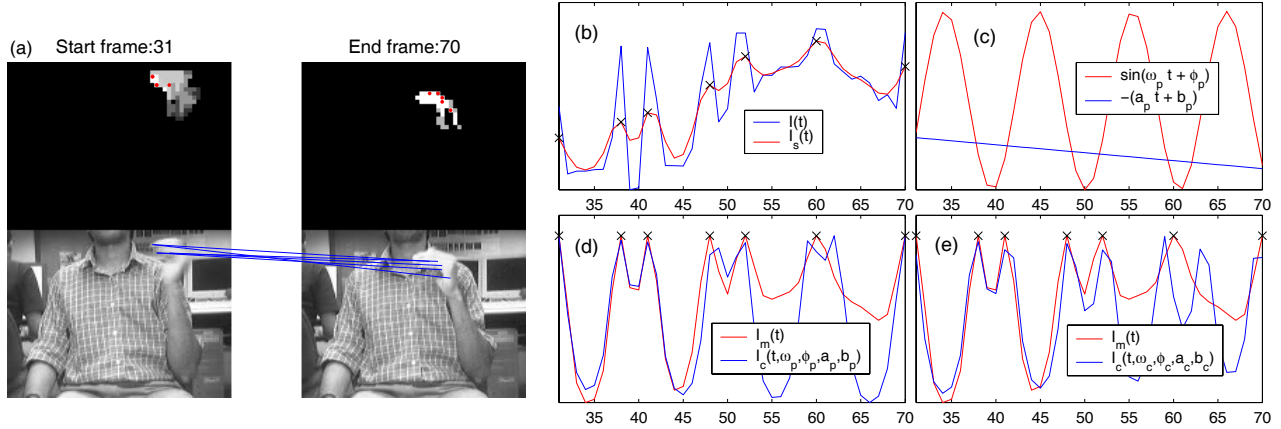
Figure 2: (a) Motion estimate (after thresholding) for boundary frames of a video segment and sample points drawn from high motion-magnitude pixels and linear sample paths connecting *all pairs* of sample points in space-time (sample point correspondence between the boundary frames is not needed). (b) The intensity profile $I(t)$ corresponding to sample path in (a). Note the peaks at locations where the hand intersects the sample path. The smoothed profile $I_s(t)$ with detected peaks is plotted. (c) $(\omega_p, \phi_p, a_p, b_p)$ are parameters estimated using only the peak locations (Sec. 3.3). (d) $I_m(t)$ is the result of setting peaks in $I_s(t)$ to 1. $I_c(t, \omega_p, \phi_p, a_p, b_p)$ (Eqn. 4) is the continuous solution at the peak estimate. (e) The solution $I_c(t, \omega_c, \phi_c, a_c, b_c)$ with parameters estimated using the continuous formulation (Sec. 3.4).

3.3. Least squares solution using intensity peaks

Task: Given the location of intensity peaks in time $\mathbf{t^P} = [t_0, t_1, \ldots t_m]^T$ along a sample path, estimate the object motion and sample path parameters.

The error vector $\mathbf{e^P}$ and the squared error E_p for an estimate (ω, ϕ, a, b) are given by,

$$e_i^p(\omega, \phi, a, b) = \sin(\omega t_i^p + \phi) + a t_i^p + b \quad (1)$$

$$E_p(\omega, \phi, a, b) = \sum_{i=1}^{m} (e_i^p)^2. \quad (2)$$

Note that we have inverted the signs of a, b for convenience. We can now formulate an unconstrained least squares minimization problem as follows,

$$\min_{\omega, \phi, a, b} E_p \quad (3)$$

$$\nabla_\omega E_p = 2 \sum_{i=1}^{m} \cos(\omega t_i^p + \phi) \, t_i^p \, e_i^p$$
$$\nabla_\phi E_p = 2 \sum_{i=1}^{m} \cos(\omega t_i^p + \phi) \, e_i^p$$
$$\nabla_a E_p = 2 \sum_{i=1}^{m} t_i^p \, e_i^p \qquad \nabla_b E_p = 2 \sum_{i=1}^{m} e_i^p.$$

To apply a gradient descent based minimization technique, we first need to analyze the local minima behavior of E_p. Consider varying each parameter keeping others fixed. Clearly the objective function $E_p(\omega, \phi, a, b)$ is quadratic in a, b and behaves as $sin^2(\phi)$ in ϕ. Hence ω is the only parameter with possible multiple local minima. We can easily handle this by choosing multiple starting points with uniformly spaced samples for $\omega \in [\omega_{\min}, \omega_{\max}]$ and choosing the global minimum. Note that since we are sampling only one parameter, the global optimization is very efficient. We use the Matlab function *fminunc* with gradients enabled to

solve this. Fig. 2(b) shows an example of the solution obtained.

3.4. Least squares solution using all intensity values

Task: Estimate the object motion and sample path parameters using all intensity values $I(t)$ for a sample path.

Merely using peak locations is not sufficient in practice due to noise and blurring effects of large or slowly moving objects. We hence wish to formulate a new objective that uses all the information available in the intensity samples. We assume that the intensity profile can be approximated by the following function,

$$I_c(t) = \frac{1}{1 + C(\sin(\omega t + \phi) + at + b)^2} \quad (4)$$

$I_c(t)$ is a function that has peaks of 1 at t values where $\sin(\omega t + \phi) + at + b = 0$. C is a constant which controls the width of the peaks. Large C values give narrow peaks approximating small objects or fast moving objects, i.e. with large motion amplitude. Small C values give broad peaks approximating large objects or slowly moving objects. It is also possible to include C as a parameter in the optimization but in the interest of efficiency we choose an empirically good C value in the range $[1.5 - 2]$. Fig. 2(d) plots $I_c(t)$ evaluated at parameters obtained using the peak locations.

We wish to formulate a least squares solution to minimize $I_c(t) - I(t) \, \forall t$, where $I(t)$ are the actual intensity values along a sample path. Before proceeding with the minimization, we need to normalize the actual intensity profile, $I(t)$ to look similar to that of the assumed profile, $I_c(t)$.

Smoothing and normalization of intensity: To alleviate noise, $I(t)$ is filtered with a Gaussian kernel ($\sigma = 1.1$) and

scaled to $[0, 1]$ to obtain $I_s(t)$, a smooth intensity profile. The peak intensity values of $I_s(t)$ are not necessarily at 1 and are typically distributed over a wide range of values. This would hence lead to a bad fit during minimization. To overcome this, we pull the peak intensity values in $I_s(t)$ to 1 using radial basis function (RBF) interpolation [8] to get a modified intensity profile, $I_m(t)$. We use RBFs since the width of influence is easily controlled by changing σ of the Gaussian kernel (we use $\sigma = 1$). Fig. 2(b) plots an example smoothed intensity, $I_s(t)$ and (d) plots the corresponding renormalized intensity, $I_m(t)$. We now use $I_m(t)$ instead of $I(t)$ in a least squares formulation to estimate parameters.

We define an error vector $e_i^c(\omega, \phi, a, b) = I_c(t_i) - I_m(t_i)$ and squared error $E_c(\omega, \phi, a, b) = \sum_{i=1}^{n}(e_i^c)^2$. Here $t = [t_1, t_2...t_n] = [1, 2...n]$, the sampling time values. The minimization of E_c can hence be formulated as,

$$\min_{\omega, \phi, a, b} E_c \qquad (5)$$

$$\text{Let } d_i = -2C \left[1 + C(\sin(\omega t_i + \phi) + at_i + b)^2\right]^{-2}$$
$$\times (\sin(\omega t_i + \phi) + at_i + b)$$
$$\nabla_\omega E_c = \sum_{i=1}^{n} d_i \cos(\omega t_i + \phi)t_i e_i^c$$
$$\nabla_\phi E_c = \sum_{i=1}^{n} d_i \cos(\omega t_i + \phi)e_i^c$$
$$\nabla_a E_c = \sum_{i=1}^{n} d_i t_i e_i^c, \qquad \nabla_b E_c = \sum_{i=1}^{n} d_i e_i^c.$$

In this case, E_c no longer has the nice quadratic properties seen with E_p in Eqn. 3. We can instead have local minima in potentially each of the parameters. This necessitates the choice of a good initialization point for minimization. We use the estimate obtained using only the peak locations (Eqn. 3) as a starting point. In practice we see that this is a good choice and the optimization quickly converges to the true value. Fig. 2(e) displays $I_c(t)$ evaluated at the solution of the above continuous formulation.

3.5. Consensus among sample paths and re-sampling new paths

We employ a RANSAC-type approach to obtain a robust period estimate. We run the least squares estimation procedure independently on the intensity profiles for all sample paths. The period estimates from many paths are binned to form a period-histogram or periodogram. The highest frequency bin is chosen as the candidate period. We extract a second set of sample paths oriented close to the candidate paths. We achieve this by connecting randomly chosen sample points inside a 5×5 window at both ends of the candidate path. We re-run the estimation process to get a new histogram. In practice, we have observed that re-sampling once is sufficient to obtain an accurate and robust estimate. This process also helps to weed out sample paths that sample from two or more objects.
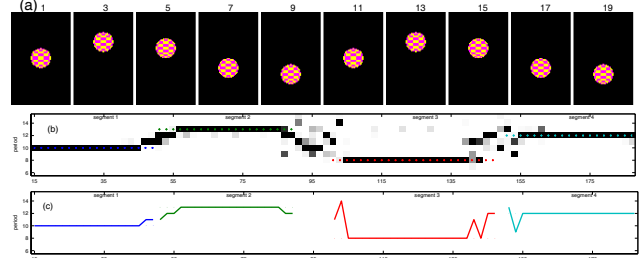

Figure 3: (a) Frames from a synthetic video of a reciprocating checkered sphere. (b) Estimated periodogram for synthetic video, the x axis corresponds to the frame number, the y axis represents periods. Each column shows the period histogram estimated from a window of 30 frames, dark blocks correspond to peaks in the histogram. The dots in each column represent the ground truth period values. (c) Plots the peaks in the periodogram. The estimated values match the true values except in transition regions.

4. Experiments

In our discussion below, we measure the period in number of frames. We first validate our technique on a simple synthetic video sequence generated by a reciprocating checkered sphere, Fig. 3(a). This video consists of periodic segments with different periods (10, 13, 8 and 12), the length of each periodic segment is 50 frames. The estimated periodogram using a window of length 30 frames is displayed in Fig. 3(b). Clearly, the estimated periods are accurate (peaks in the histogram correspond to the true values) except in transition regions. This validates that the proposed technique correctly estimates the right period for different orientations of sample paths with respect to axis of object motion. We tried similar experiments changing the size of sphere and amplitude of motion and obtained accurate results. In these cases we need to appropriately set the C value in Eqn. 4 since the width of peaks in intensity change with object size and motion amplitude, the C values we used were in the range $[1.5 - 2]$.

4.1. Experiments on real video

We use grayscale video sequences of size 160×120 captured at 15 fps of subjects performing the same set of periodic motions in succession. Each video has nine contiguous *periodic motion segments* of different cyclic signs as displayed in Fig. 4. Each segment captures $5 - 10$ cycles of a sign. We used seven subjects in our experiments. In our discussion below, we use *frame window* or *window* to mean a fixed length window of frames used in estimation. We present results of periodicity estimation using the proposed technique with frame window sizes of 30 and 40. These sizes were chosen since we need at least two periods in each window to obtain a robust estimate. We extract 100 sample paths in each window to obtain the first estimate. We then re-sample 25 paths close to the candidate paths as described in Sec. 3.5. Our Matlab implementation takes about
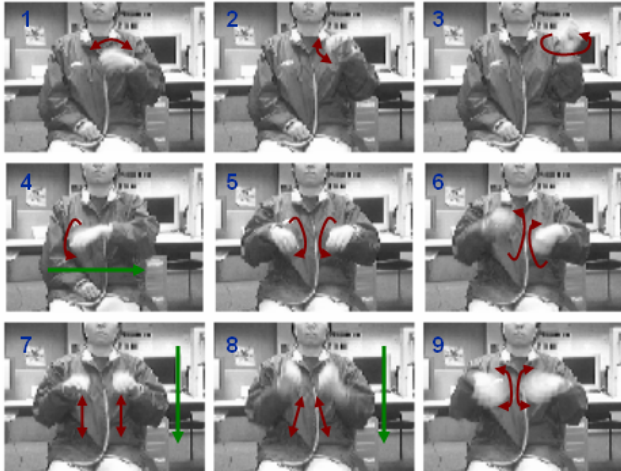
Figure 4: A real video sequence showing nine types of periodic motions used in our experiments. Red arrows show the periodic motion of the hands. Green arrows show direction of global motion for segments 4,7 and 8.

two minutes on a 2GHz, 64bit-AMD processor to estimate parameters for all sample paths with re-sampling for each frame window. Fig. 5 plots the periodograms after one iteration of re-sampling.

4.2. Performance evaluation

To evaluate the performance of the proposed technique, we obtain ground truth by hand-labeling all cycles in each video. Each frame window typically overlaps $3 - 5$ cycles of a sign. Even within the same periodic motion segment, the period of human motion varies between adjacent cycles. We hence have a range of true period values for each frame window. We display this variability in Fig. 5 using three dots representing the min, avg, and max values of true period in the corresponding frame window. In Fig. 5 we plot peaks in the histogram for successive frame windows along with the ground truth, vertical lines display the min and max true period values in each frame window.

Measurement of estimation error: We compute the error in period estimate for each frame window as follows. If the estimated period given by the peak of the periodogram lies within the true range, we do not count it as an error. If the estimate is beyond the true range, we use the difference from the closest end of the range as the error value. We only measure error in frame windows that correspond to periodic motion segments in the ground truth. Table 1 shows the average error for each periodic motion segment with window lengths of 30 and 40 frames.

In cases where the ground truth is good, i.e. the variability in period is low; e.g. subject v – segments $3, 6$, subject k – segments $3, 6$, subject q – segments $2, 3, 4, 5, 6$ in Fig. 5; our techinque estimates the correct period value. We observe that the period estimates for one handed signs

(columns $1 - 5$) are more accurate then estimates for two handed signs (columns $7 - 9$) in Table 1. The subjects were able to perform uniform periodic motions for one handed signs, but the variability in period for most subjects with two handed signs was higher leading to many different periods in a frame window. This in turn leads to performance deterioration.

In some cases with two handed signs, a majority of the initially chosen paths sample from both hands and hence our technique is unable to recover the correct period. It is interesting to observe that in the case of the two handed signs in columns $5, 6$ in Fig. 5 (periodic motions shown in Fig. 4) we see good performance comparable to that of one handed signs. This is because most subjects were able perform in-phase and out of phase rotation of both hands at a constant rate (this can be seen from the low variability of ground truth periods in columns $5, 6$).

Comparing the results for different window sizes, we see that the average error in estimate is smaller for most users and periodic motion segments with window size of 30. This is again due to the fact that with longer window size the variability in the data increases, also non-linearity of motion affects performance.

5. Summary and Conclusions

In contrast to prior work, our method requires neither feature tracking nor object segmentation. The intensity patterns along linear sample paths in space-time is used in estimation of period of object motion in a given sequence of frames. We show through experiments with low resolution video for a wide variety of cyclic signs, the proposed technique on average estimates period values within 1 frame. Possible applications of our technique include, segmenting video streams into "periodic clips" for annotating ASL, using detected periodic motion to recover regions of pixels that are consistent with the periodic motion; applications in pedestrian detection, as well as separating individuals walking in groups.

As part of future work, we plan to extend our formulation to handle sampling from multiple objects. The number of repetitions of signs in typical sign language video is less than three and hence our technique is unable to recover the correct period. Considering that humans can easily identify such repetitions, a robust learning component is needed to achieve similar performance. We are also exploring techniques to reorient sample paths in space-time so as to sample from objects at regular intervals.

References

[1] S. Niyogi and E. Adelson, "Analyzing and Recognizing Walking Figures in XYT," *Proc. CVPR*, pp. 469-474, 1994.

Sub-	Average error in each video segment									Row
ject	1	2	3	4	5	6	7	8	9	Mean
a	0.4	0.3	0.4	0.4	0.7	0.8	0.4	1.3	0.2	0.5
k	0.2	0.1	0.2	0.4	0.2	0.0	0.0	1.1	0.5	0.3
v	0.0	0.2	0.0	0.0	0.1	0.0	0.2	1.0	0.5	0.2
t	0.3	0.3	0.3	1.0	0.6	0.6	0.0	0.5	2.2	0.6
r	0.0	0.6	0.3	0.1	0.9	0.1	0.6	1.8	1.3	0.6
g	0.7	0.9	0.8	0.6	0.2	0.2	2.6	2.3	1.0	1.0
q	0.5	0.5	0.2	0.0	0.2	0.2	0.8	2.1	0.3	0.5
Mean	0.3	0.4	0.3	0.4	0.4	0.3	0.7	1.4	0.9	0.5

(a) window size 30 frames

Sub-	Average error in each video segment									Row
ject	1	2	3	4	5	6	7	8	9	Mean
a	0.0	0.3	0.1	0.0	0.0	0.1	0.4	2.3	0.4	0.4
k	0.1	0.0	0.0	0.1	0.1	0.0	0.1	2.0	0.5	0.3
v	0.2	0.3	0.0	0.1	0.2	0.2	0.1	1.2	0.8	0.3
t	0.0	0.5	0.4	1.2	0.6	0.0	0.3	0.4	2.8	0.7
r	0.0	1.2	0.0	0.5	1.0	0.0	0.7	0.3	2.1	0.7
g	0.4	1.8	0.8	2.6	0.4	0.2	1.2	3.5	2.2	1.5
q	1.8	1.3	0.0	0.0	0.9	0.1	1.3	1.6	0.4	0.8
Mean	0.4	0.8	0.2	0.6	0.5	0.1	0.6	1.6	1.3	0.6

(b) window size 40 frames

Table 1: Average errors in estimated period values for seven subjects are shown. Each video segment corresponds to a periodic motion shown in Fig. 4. The error in estimated period values averaged over all subjects is 0.5 frames for a window length of 30 and 0.6 frames for a window length of 40. The range of true period values in these videos was $5 - 14$ frames.
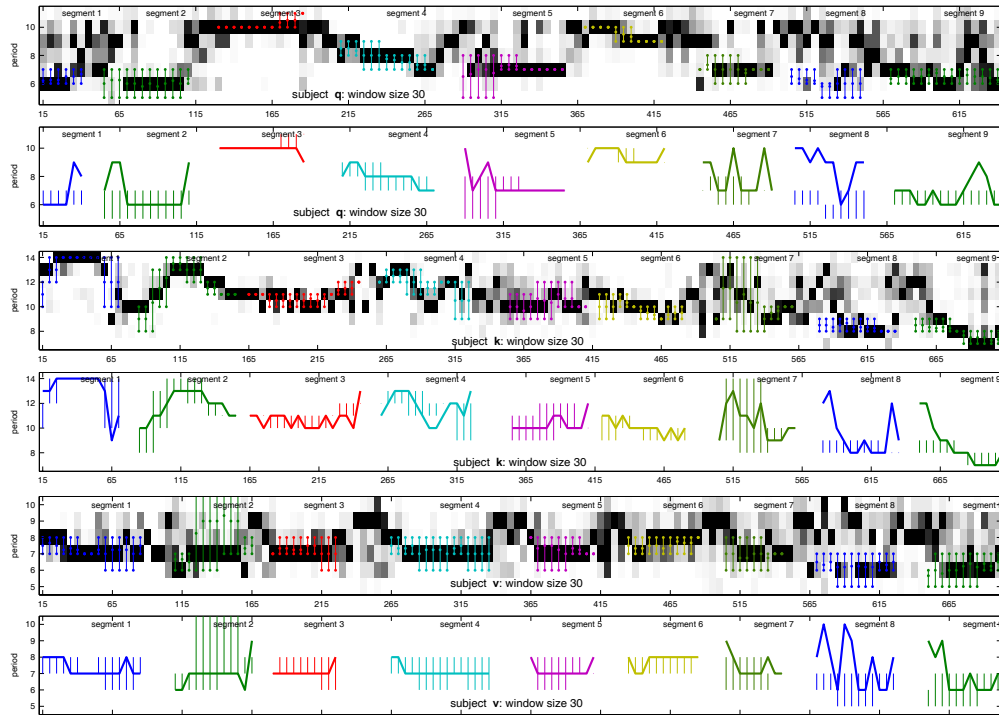

Figure 5: The estimated periodograms for three subjects are shown. Each column in the periodogram represents the histogram of estimated period values for a frame window of length 30 frames from 125 sample paths. Dark blocks in the image correspond to peaks in the histogram. The range of ground truth period values are shown as three dots in each column and correspond to the min, avg and max values of true period. Below each periodogram, we plot peaks in estimated period histogram as thick lines and the range of *true period values* in each frame window as thin vertical lines.

[2] F. Liu and R. Picard, "Finding Periodicity in Space and Time," *Proc. ICCV*, pp. 376-383, 1998.

[3] S.M. Seitz and C.R. Dyer, "View-Invariant Analysis of Cyclic Motion," *IJCV*, 25(3), pp. 1-23, 1997.

[4] R. Cutler and L.S. Davis, "Robust Real-Time Periodic Motion Detection, Analysis, and Applications," *T-PAMI*, 22(8), pp. 781-796, 2000.

[5] D. Ormoneit and H. Sidenbladh and M. J. Black and T. Hastie and D. J. Fleet, " Learning and Tracking Human Motion Using Functional Analysis,",*Proc. IEEE Workshop on Human Modeling, Analysis and Synthesis*, 2000.

[6] J. Davis and A. Bobick and W. Richards, "Categroical Represntation and Recognition of Oscillatory Motion Patterns,",*Proc. CVPR*, 2000.

[7] Y. Xiong and F. Quek and D. McNeill, "Hand Motion Gestural Oscillations Multimodal Discourse," *Proc. Intl. Conf. on Multimodal Interfaces*, 2003.

[8] S. Haykin, "Neural Networks: A Comprehensive Foundation," *Prentice Hall Intl.*, pp. 262-264, 1999.

[9] L. Sigal and S. Sclaroff and V. Athitsos, "Skin color-based video segmentation under time-varying illumination," *T-PAMI*, 26(7), pp. 862-877, 2004.