Estimation of Multiple Periodic Motions from Video

Alexia Briassouli and Narendra Ahuja

Beckman Insitute, University of Illinois, Urbana-Champaign,
405 N Matthews, Urbana, IL, 61801
{briassou, ahuja}@vision.ai.uiuc.edu

Abstract. The analysis of periodic or repetitive motions is useful in many applications, both in the natural and the man-made world. An important example is the recognition of human and animal activities. Existing methods for the analysis of periodic motions first extract motion trajectories, e.g. via correlation, or feature point matching. We present a new approach, which takes advantage of both the frequency and spatial information of the video. The 2D spatial Fourier transform is applied to each frame, and time-frequency distributions are then used to estimate the time-varying object motions. Thus, multiple periodic trajectories are extracted and their periods are estimated. The period information is finally used to segment the periodically moving objects. Unlike existing methods, our approach estimates multiple periodicities simultaneously, it is robust to deviations from strictly periodic motion, and estimates periodicities superposed on translations. Experiments with synthetic and real sequences display the capabilities and limitations of this approach. Supplementary material is provided, showing the video sequences used in the experiments.

1 Introduction

Periodic motion characterizes the motion of humans and animals, as well as many man-made objects [1]. This paper presents a new approach to the analysis of multiple periodic motions in a video sequence. The primary motivation and intuition lie in the observation that repetitive patterns have distinct frequency space signatures. If these signatures can be extracted, then they can be used to enhance the more common, spatial domain analysis of the video sequence. This synergy between periodic motion and frequency space representations has been surprisingly underexploited.

The main parts of the proposed approach are as follows. (1) Through a process called μ-propagation, the periodic changes in object motions are converted into a proportional variation in frequency (Sec. 4). This results in a frequency-modulated (FM) signal with time-varying frequencies. (2) Time-frequency distributions (TFDs) are used to estimate the time-varying frequencies, and the periods present in them are estimated via spectral analysis methods (Sec. 3, 4). (3) Once all the periods in the video sequence are estimated, each object is

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part I, LNCS 3951, pp. 147–159, 2006.

segmented (Sec. 5) by matching each frame with frames at displacements corresponding to its period (since an object is expected to re-appear in the same position after an integer number of periods).

1.1 Previous Work

The numerous methods for analyzing repetitive motions can be separated in two large categories: the first based on the analysis of feature correspondences, and the second category on region correlations.

Point Correspondence Methods: Much of the work on periodic motion estimation and analysis [2], [3] extracts the trajectories by tracking the position of reflective markers throughout the video. When manual intervention or the placement of markers are not possible, feature correspondences are used. However, varying illumination, or local occlusion lead to point feature detection and localization errors, making the point matching unreliable. Given the detected point features in each image, the large numbers of possible pairings also make them computationally forbidding for many applications.

Region Correspondences: Region based methods [4] find repetitions in interframe region correlations [5]. They avoid the sensitivity of point correspondences, but are still sensitive to non-constant illumination. Also, they detect "in position" periodicities, i.e. oscillating positions of the objects around the same pixel(s). They cannot detect periodicities superposed on other motions, such as translations (e.g. walking), without pre-processing. Pre-processing requires that each oscillating object is segmented in each frame [4], [6] and then aligned in successive frames, to detect periodicities.

1.2 Motivation

The proposed work is strongly motivated by the aforementioned frequency-compatible nature of periodic motion analysis, the limitations of the current, spatially based methods, and the potential advantages of combining the strengths of spatial and frequency based approaches. The advantages the frequency based methods [7], [8] introduce include the following. (1) Frequency-based approaches involve spatially global, instead of local, analysis. (2) There is no need for explicit feature matching (as in spatial methods). (3) Frequency domain analysis is robust to illumination changes: Fourier Transform (FT) based motion estimates are extracted from phase changes induced by motions, which are not as sensitive to illumination changes as spatial correlations [9]. (4) Efficient algorithms are available for FT computation.

1.3 Contributions

The major contributions of the proposed approach are: (1) Unlike previous work, it extracts multiple periodic motions. (2) Periodic trajectories are extracted simultaneously, not one at a time (Sec. 5). (3) It is robust to deviations from strict periodicity (Sec. 6). For example, (a) when the period is not truly constant, or

(b) when the magnitude of the velocity or displacement profile does not have the exact same value at each repetition, or (c) when object shape is not rigid, and all or some of the motion parameters fluctuate around some "mean" values, the effects of these deviations on the proposed approach are marginal (Sec. 9). (4) The computational cost is lower than that of the spatial methods, because (a) the FT computation is efficient, and (b) frame by frame processing is reduced to a few frame correlations for segmentation (Step (3) in Sec. 1). (5) It is an example for formulating joint spatial and frequency solutions to other problems.

2 Mathematical Formulation

Consider M periodically moving objects $s_i(\bar{r})$, $1 \leq i \leq M$, with no interobject occlusion, and a still background $s_b(\bar{r})$. In the spatial domain, frame 1 is $a(\bar{r}, 1) = s_b(\bar{r}) + \sum_{i=1}^{M} s_i(\bar{r}) + v_{noise}(\bar{r}, 1)$. The objects actually mask background areas [10], so a more accurate model is acquired by removing (setting to 0) the background in each frame[1]. Then, frame n $(1 \leq n \leq N)$ is $a(x, y, n) = \sum_{i=1}^{M} s_i(x - b_i^x(n), y - b_i^y(n)) + v_{noise}(x, y, n)$, where $\bar{b}_i(n) = [b_i^x(n), b_i^y(n)]$ represents the displacement of object i, $1 \leq i \leq M$ from frame 1 to n, $1 \leq n \leq N$. Its 2D FT is:

$$A(\omega_x, \omega_y, n) = \sum_{i=1}^{M} S_i(\omega_x, \omega_y) e^{-j(\omega_x b_i^x(n) + \omega_y b_i^y(n))} + V_{noise}(\omega_x, \omega_y, n). \quad (1)$$

$A(\omega_x, \omega_y, n)$ has $b_i^x(n)$ and $b_i^y(n)$ as linear terms in its phase, and consequently it has a time-varying spectrum. The latter cannot be estimated via the 3D FFT, since the motion is not constant, as in [11]. Alternate methods are needed if we wish to estimate the periodicity in b_i^x and b_i^y from the spectral variations.

3 Short Term Fourier Transform

Non-stationary signals, i.e. signals with time-varying spectra, can be analyzed with time-frequency distributions (TFD's), which capture the variations of the frequency content of the signal with time [7]. We use the Short-Term Fourier Transform (STFT), which is the most common TFD [12]. The STFT captures the spectral variation with time by computing the FT of the local signal, by filtering it with an appropriate low-pass time function. The spectrum of the filtered signal represents the spectral content of the signal at that time instant. For a 1D signal $s(t)$, the STFT is defined as $STFT_s(t, \omega; h) \equiv \int_{-\infty}^{+\infty} s(\tau + t) h^*(\tau) e^{-j\omega\tau} d\tau$, where $h(t)$ is a lowpass function representing the "analysis window". There is an inherent tradeoff between time and frequency resolutions, depending on the window used: if $h(t)$ has higher values near the observation point t, the STFT estimates more local quantities. A window that is compact in time leads to higher

[1] In general, the background at each pixel can be estimated from the observed intensity distributions at each pixel, and its recognition as background will involve a statistical decision. We will omit the details of this step in this paper.

time resolution, whereas a window peaked in the frequency domain gives better frequency resolution.

4 Time-Varying Frequency Estimation

The time-varying frequency of the signal $A(\omega_x, \omega_y, n)$ in Eq. (1) can be estimated by applying the TFDs, which have been used for 1D signals [13]. They have also been used for motion estimation [14], but for horizontal or vertical projections of the video, i.e. 1D signals again. Here, we present a method that can estimate the 2D object motions without resorting to projections.

Consider frame $a(x, y, n)$. We construct an FM signal, whose 2D frequency is modulated by the time-varying displacements of the objects, via constant μ propagation [14]. Essentially, we estimate the 2D FT at a constant 2D "spatial frequency" $\bar{\mu} = [\mu_1, \mu_2]$, as follows:

$$A(\mu_1, \mu_2, n) = \sum_x \sum_y \sum_{i=1}^{M} [s_i(x - b_i^x(n), y - b_i^y(n)) + v_{noise}(x, y, n)] e^{j(\mu_1 x + \mu_2 y)}$$

$$= \sum_{i=1}^{M} S_i(\mu_1, \mu_2) e^{j\mu_1 b_i^x(n)} e^{j\mu_2 b_i^y(n)} + V_{noise}(\mu_1, \mu_2).$$

The frequencies $\omega_i(n) = \mu_1 b_i^x(n) + \mu_2 b_i^y(n)$ in $A(\mu_1, \mu_2, n)$ are extracted by applying TFDs to that signal. However, the motion appears in each $\omega_i(n)$ as a weighted sum of the horizontal and vertical displacements. This problem can be overcome simply, by estimating $A(\mu_1, \mu_2, n)$ at $\mu_1 = 0$ and $\mu_2 = 0$. This gives $\omega_i(n) = \mu_2 b_i^y(n)$ and $\omega_i(n) = \mu_1 b_i^x(n)$ respectively, so the horizontal and vertical displacements are separated.

Using TFD's, the multiple frequencies are represented by multiple ridges in the time-frequency plane, which show the power spectrum corresponding to each time and frequency instant. The peaks of these ridges give the dominant frequencies at each time n, leading to a multicomponent signal, consisting of the M time-varying frequencies $\omega_i(n)$, one for each object $1 \le i \le M$.

5 Multiple Period Detection and Estimation

We introduce a simple but efficient method for the recovery of the M different repetitive components of the object motions, that takes advantage of their periodic nature. At each frame n, we have M displacement values $b_1^x(n), ..., b_M^x(n)$ and $b_1^y(n), ..., b_M^y(n)$. For each object, the $b_i^x(n)$, $b_i^y(n)$ form periodic functions of time. We examine only the horizontal trajectories, since the same analysis can be applied to the vertical ones. For object i, $1 \le i \le M$, and time n, $1 \le n \le N$, we get the periodic signal $\bar{b}_i^x = [b_i^x(1), ..., b_i^x(N)]$, representing its motion over time. We sum the M signals $\bar{b}_i^x$ of all objects i at each instant n, to form the function $\bar{g}_x = [g^x(1), ..., g^x(N)] = \sum_{i=1}^{M} \bar{b}_i^x$, with values at each frame n $(1 \le n \le N)$

given by $g^x(n) = \sum_{i=1}^{M} b_i^x(n)$. The resulting 1D function $\bar{g}_x$ is a sum of periodic functions $\bar{b}_i^x$, with different periods T_i^x $(1 \le i \le M)$. Traditional spectral analysis methods (e.g. the MUSIC algorithm) give the M frequencies ω_i^x $(1 \le i \le M)$ of $\bar{g}_x$, and the corresponding periods $T_i^x = 1/\omega_i^x$. The details of the spectral analysis methods used are omitted, as they are beyond the scope of this paper, and well documented in the literature [15], [8].

5.1 Periodically Moving Object Extraction

Once the different periods are estimated, the moving objects can also be extracted: by correlating frames separated by an integer number of periods, we expect to get higher correlation values in the area of each periodically moving object. We have $b_i^x(n) = b_i^x(n + T_i^x)$, $b_i^y(n) = b_i^y(n + T_i^y)$ for object i. We consider $T_i^x = T_i^y = T_i$ for simplicity, but the same analysis can be applied when $T_i^x \neq T_i^y$. If T_j denotes the period of object j, at time $n' = n + T_j$ we have:

$$a(x, y, n') = \sum_{i=1}^{M} s_i(x - b_i^x(n'), y - b_i^y(n')) + v_{noise}(x, y, n')$$

$$= \sum_{i \neq j} s_i(x - b_i^x(n'), y - b_i^y(n')) + s_j(x - b_j^x(n'), y - b_j^y(n')) + v_{noise}(x, y, n')$$

since object j is in the same position in frames n and $n' = n + T_j$. Therefore, we can extract the j_{th} object by correlating frames n and $n' = n + T_j$: since only that object is expected to re-appear in the same position in those frames, the correlation values will be highest in the pixels in its area.

5.2 Object Extraction for Periodic Motion Superposed on Translation

As stated in Sec. 1, one of the main contributions of our method is the fact that it allows the estimation of periodic motions superposed on translations, such as walking. In these cases, the legs are moving periodically, but the moving entity is also translating. Correlation-based methods cannot deal with such motions, because of the shifting position of the periodically moving object. The time-varying trajectory $b(n)$, which is used to create the FM signal, is of the form $b(n) = \alpha \cdot n + b_P(n)$, where $1 \le n \le N$, α is a constant and $b_P(n)$ is the periodic component of the motion. The FM signal we create via μ-propagation is $z(n) = e^{j\mu(\alpha \cdot n + b_P(n))}$, with phase $\phi_z(n) = \mu(\alpha \cdot n + b_P(n))$. The TFDs estimate its frequency, i.e. the time-derivative of $\phi_z(n)$, $\omega_z(n) = \frac{\partial(j\mu(\alpha \cdot n + b_P(n)))}{\partial n} = j\mu\alpha n + \frac{\partial b_P(n)}{\partial n}$. Consequently, the translational component of the motion becomes a simple additive term, whereas the periodicity of $b_P(n)$ is retained in the extracted frequency. This allows us to deal with periodic motions superposed on translations, without needing to align the video frames.

The segmentation cannot be performed directly in terms of the periodic motion parameters, since the object has also translated. This difficulty can be easily overcome by estimating the “mean” translation between frames, via their

FT [9], [10]. If there are M objects in the sequence, where object i is displaced by $\bar{b}_i(n)$ from frame 1 to n, the ratio of the FTs of frame n (Eq. (1)) and frame 1 is $\phi_n(\bar{\omega}) = \frac{A(\bar{\omega},n)}{A(\bar{\omega},1)} = \sum_{i=1}^{M} \gamma_i(\bar{\omega}) e^{-j\bar{\omega}^T \bar{b}_i(n)} + \gamma_n(\bar{\omega})$, where $\gamma_i(\bar{\omega}) = \frac{S_i(\bar{\omega})}{A(\bar{\omega},1)}$, $\gamma_n(\bar{\omega},n) = \frac{V_{noise}(\bar{\omega},n)}{A_1(\bar{\omega})}$. Its inverse FT is:

$$\phi_n(\bar{r}) = \sum_{i=1}^{M} \gamma_i(\bar{r}) \delta(\bar{r} - \bar{b}_i(n)) + \gamma_n(\bar{r}, n), \tag{2}$$

so it has peaks at $\bar{r} = \bar{b}_i(n)$, for $1 \leq i \leq M$. Thus, the peaks of $\phi_n(\bar{r})$ estimate the "mean" translations $\bar{b}_i(n)$ of object centroids, between frames 1 and n.

6 Evaluation of the Robustness of the Estimates

Although many motions appearing in nature and in man-made applications have a repetitive form, they are not necessarily strictly periodic. In most cases, their period may fluctuate around a "mean period", and the peak displacement may exhibit similar deviations around a mean value. For the analysis here, we consider one object, and only the motion in the x-direction since the same applies to the y-direction. Consider an ideal periodic trajectory $x(t) = x(t+T)$, and a nearly periodic trajectory $x'(t) = x(t+T') + \epsilon_2$, where $T' = T + \epsilon_1$, $\epsilon_1 \sim \mathcal{N}(0, \sigma_1^2)$, $\epsilon_2 \sim \mathcal{N}(0, \sigma_2^2)$. The analysis will be carried out in continuous time, so the signal under examination is $A(\mu_1, 0, t) = S(\mu_1, 0, t) e^{j\mu_1 x(t)}$, with STFT $STFT'(t,\omega) = \int S(\mu_1, 0) e^{j\mu_1 x(t+\tau)} h^*(\tau) e^{-j\omega\tau} d\tau$. For a near-periodic trajectory $x'(t)$, the STFT is $STFT(t,\omega) = \int S(\mu_1, 0) e^{j\mu_1 (x(t+\tau+T+\epsilon_1)+\epsilon_2)} h^*(\tau) e^{-j\omega\tau} d\tau$. The noise in the displacement period and peak magnitude introduce errors in the STFT, which is a random quantity. Its mean, w.r.t. the random quantities ϵ_1, ϵ_2, is $E_{\epsilon_1,\epsilon_2}[STFT'(t,\omega)] = E_{\epsilon_1} E_{\epsilon_2}[STFT'(t,\omega)] = E_{\epsilon_2}[e^{j\mu_1\epsilon_2}] E_{\epsilon_1}[F(\epsilon_1)]$, where $F(\epsilon_1) = S(\mu_1, 0) \int e^{j\mu_1 x(t+\tau+T+\epsilon_1)} h^*(\tau) e^{-j\omega\tau} d\tau$. Then:

$$E_{\epsilon_2}[e^{j\mu_1\epsilon_2}] = \frac{1}{\sqrt{2\pi}\sigma_2} \int_{-\Delta_2}^{\Delta_2} exp\left[-\frac{1}{2}\left(\frac{\epsilon_2^2}{\sigma_2^2} - 2j\mu_1\epsilon_2\right)\right] d\epsilon_2. \tag{3}$$

For $z = \frac{\epsilon_2}{\sigma_2} - j\mu_1\sigma_2$, Eq. (3) is $E_{\epsilon_2}[e^{j\mu_1\epsilon_2}] = \frac{e^{-\frac{1}{2}\mu_1^2\sigma_2^2}}{\sqrt{2\pi}} \int_{-\Delta_2/\sigma_2 - jmu_1\sigma_2}^{\Delta_2/\sigma_2 - j\mu_1\sigma_2} e^{-z^2/2} dz$. This integral can be estimated numerically, and it can be shown that for $\sigma_2 \to 0$, $E_{\epsilon_2}[e^{j\mu_1\epsilon_2}] \to 1$. This shows that the mean STFT, with respect to the displacement magnitude error ϵ_2, is unaffected by this noise. Essentially, the STFT estimator is unbiased with respect to ϵ_2, i.e. if this error is introduced in many realizations of the trajectory, the average value of the resulting "noisy" STFTs will be the same as the true STFT. This explains why the time-frequency distribution estimate (STFT) is robust to deviations from a "perfect" trajectory, where $\epsilon_2 = 0$. For the error in the trajectory period ϵ_1, we have:

$$E_{\epsilon_1}[F(\epsilon_1)] = \frac{1}{\sqrt{2\pi}\sigma_1} \int F(\epsilon_1) e^{-\epsilon_1^2/2\sigma_1^2} d\epsilon_1 = \frac{1}{\sqrt{2\pi}\sigma_1} \int h^*(\tau) e^{-j\omega\tau} A(\tau) d\tau, \tag{4}$$

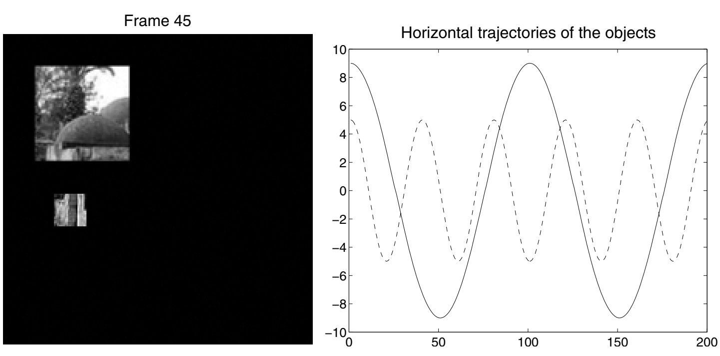
for $A(\tau) = \frac{1}{\sqrt{2\pi}\sigma_1} \int_{-\Delta_1}^{\Delta_1} e^{j\mu_1 x(t+\tau+T+\epsilon_1)} e^{-\epsilon_1^2/2\sigma_1^2} d\epsilon_1$. For $\epsilon_1 = 0$, i.e. when T is constant, Eq. (4) gives the STFT of the ideal periodic signal. $A(\tau)$ depends on the form of $x(t)$, but $E_{\epsilon_1}[F(\epsilon_1)]$ in Eq. (4) is essentially the same as the STFT of $e^{j\mu_1 x(t)}$, except after the signal $x(t)$ has been "filtered" by the Gaussian function $e^{-\epsilon_1^2/2\sigma_1^2}$. This filtering behaves like a low pass function for the signal $x(t)$, since it is blurred by the Gaussian function. Eq. (4) will give the time-frequency power spectrum of this "filtered" signal, which will lead to correct frequency estimates, since the peaks in the spectrum will simply be spread out by the blurring process.

7 Experiments

Experiments are conducted both with synthetic and real sequences that contain multiple periodic motions. Most real sequences involve only nearly periodic motions, i.e., they contain many deviations from strict periodicity. **They can be seen in the supplementary material to this paper.** The goals of the experiments are: (1) To show that the proposed method can detect multiple periodic motions. (2) To show that the multiple periods can be estimated reliably. (3) To extract the periodically moving objects.

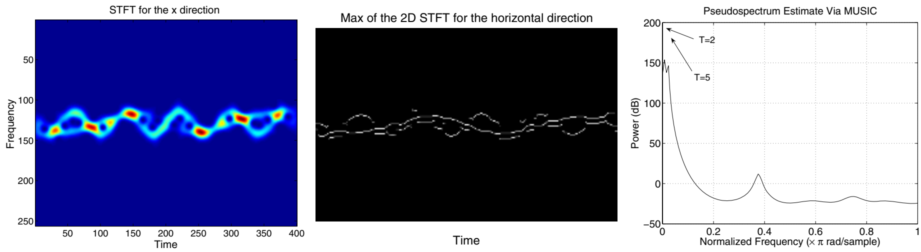
Synthetic Sequence - Two Objects: Experiments are conducted with a synthetic sequence, with horizontal motion (Fig. 1). We use μ-propagation [14] to estimate the STFT (Fig. 2(a)). The power spectrum of the STFT max (Fig. 2(b)) gives the correct periods present in the sequence (Fig. 2(c)).

Real Sequence - Walking: In this experiment we examine the case of periodic motion superposed on translation. We use the video of a person walking in parallel to the camera sensor: the human's body is translating to the left, but his legs and arms are performing repetitive motions (Fig. 3). The periods of his arms and legs are empirically found to be 5 by observing the video sequence. They are extracted correctly via the STFT, as Figs. 4 and 5 show. The mean



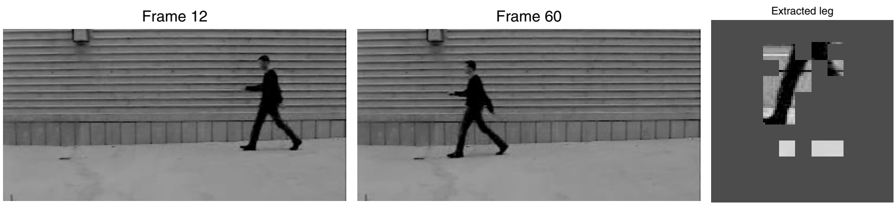
(a) (b)

Fig. 1. (a) Frame 45 of synthetic sequence with two periodically moving objects. (b) Object velocities in the horizontal direction, as functions of time.



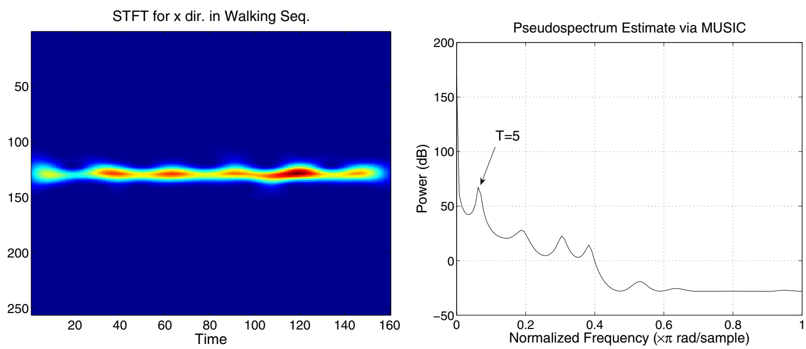
(a) (b) (c)

Fig. 2. Synthetic sequence: (a) STFT. (b) Max of the STFT. (c) The power spectrum of the TFD max gives the correct periods.



(a) (b) (c)

Fig. 3. Walking Sequence: (a) Frame 12. (b) Frame 60. (c) Segmentation of the periodically moving leg, shown in black. The deviation of the leg's motion from strict periodicity introduces blocking artifacts in the correlation process.

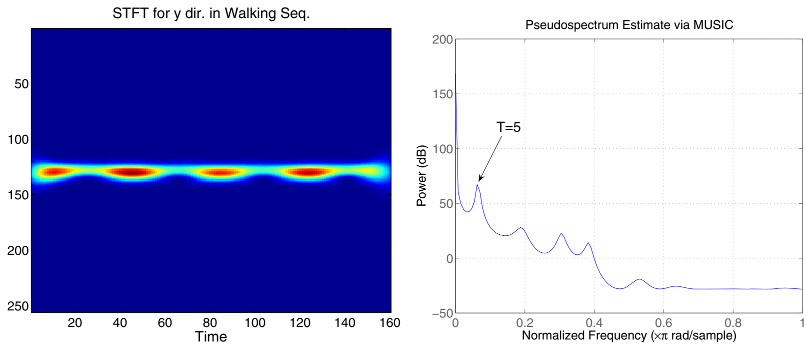


(a) (b)

Fig. 4. Horizontal direction of Walking Sequence: (a) 2D STFT (b) The power spectrum for the horizontal direction correctly finds $T = 5$ for the leg motion

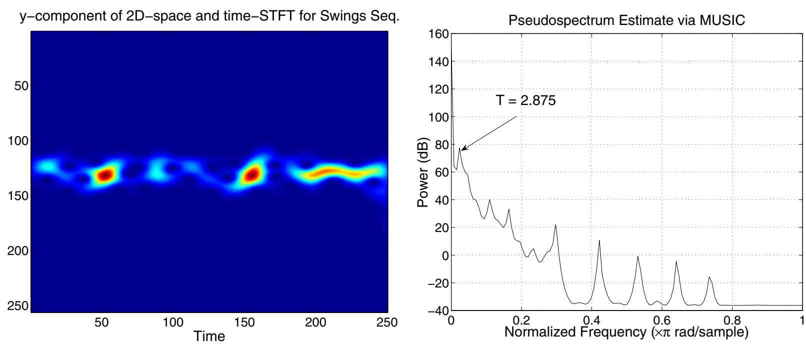
translation is then estimated to be 135 pixels via Eq. (2), and the image is shifted back to the same position in all frames. Finally, the periodically moving leg is extracted by correlating the shifted frame 60 with frame 12, corresponding to 3 periods, giving the result of Fig. 3(c).[2] In Fig. 3(c) we show only the segmented object (leg) area of the frame, shown on a larger scale than the original frames,

[2] The sequence has 80 frames and $T = 5$ so every 16 frames correspond to one period.



(a) (b)

Fig. 5. Vertical direction of Walking Sequence: (a) 2D STFT (b) The PSD for the vertical direction correctly finds $T = 5$ for the arm motion

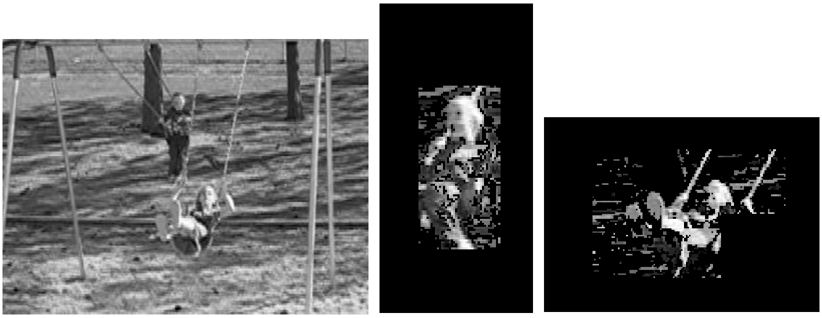


(a) (b)

Fig. 6. Swings sequence, y direction: (a) STFT. (b) Power spectrum of the STFT. The period estimate $T = 2.875$ is close to the actual value $T = 2.5$.

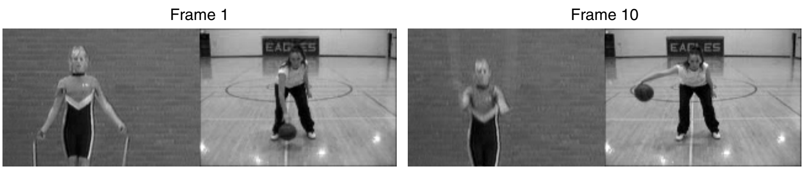
for clarity. The leg is the black part of this figure, but parts of the background have also been extracted with it during the correlation process. This is because the leg's motion is not perfectly periodic, despite its strongly repetitive nature: it is not in precisely the same position after an integer number of periods, although it is very close to its original place, as Fig. 3(a),(c) show. Thus, the correlation process also extracts some of the background around the object (leg), because of these deviations from strict periodicity.

Real Sequence - Swings: This sequence shows two children on swings (Fig. 7(a)), moving with the same period, $T = 2.5$, but different phase, as they start off from different positions. In Fig. 6(a) we see that the STFT in the y-direction captures the repetitive motions in that direction. The power spectrum of the peaks of this TFD contains the periodicity information, as shown in Fig. 6(b): the period estimate $T = 2.875$ is quite close to its observed value



(a) (b) (c)

Fig. 7. Swings sequence: (a) Frame 10. Segmentation results for (b) boy (c) girl



(a) (b)

Fig. 8. Jump-rope and dribbling sequence: (a) Frame 1 (b) Frame 10

of $T = 2.5$. It is used to correlate frames that are an integer number of periods apart, and thus segment the periodically moving children (Fig. 7(b), (c)). We show only the segmented object areas of the frame, on a larger scale than the original frames, for clarity. It should be noted that the method succeeds despite the fact that the children are non-rigid objects. Also, since they are non-rigid, the correlation is performed with large block sizes to account for the variations in their overall shape (e.g. legs folding or extending).

Real Sequence - Jump-Rope and Dribbling Sequence: In this experiment we used a sequence consisting of two different periodic motions: a girl with a jump rope, jumping in place next to a girl that is dribbling a basketball (Fig. 8). The empirically observed periods for the Jump-Rope sequence are $T_x = 4.5$ in the horizontal direction and $T_y = 8$ in the vertical direction, while in the Dribbling sequence, we have $T_x = 2.5$ and $T_y = 5$. As Fig. 9 shows, the estimated horizontal periods are $T = 2.5$ and $T = 5$, so the period of the x-movement for the dribbling is found correctly, while the jump-rope's horizontal period is estimated with a small error. This is expected, as the horizontal motion of the girl jumping is small and noisy, because of the random motion and occlusion introduced by her arms and the jump-rope. The dribbling of the ball is a more regular motion, so its period is found with better precision. Similarly, the periods of the motions in the y-direction are found to be $T = 5.4$ and $T = 7.8$ for the ball and the girl jumping, respectively. Again, they are estimated with good accuracy, although there are possible sources of errors, such as occlusion and non-rigidly moving

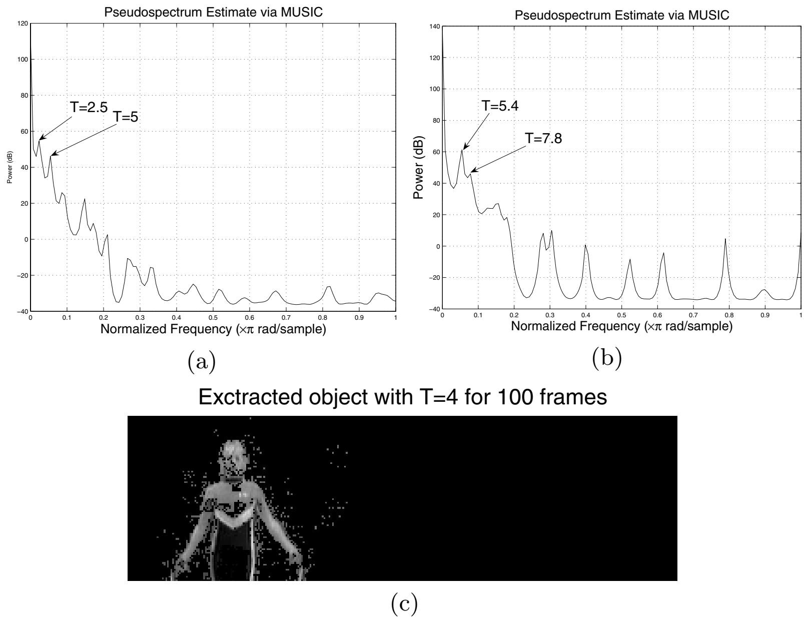


Fig. 9. Power spectral density of the 2D TFDs. (a) In the x-direction both periods are estimated correctly. (b) In the y-direction both periods are also estimated correctly. (c) The object with $T = 4$ in the x-direction extracted via spatial correlation.

objects (e.g. arms) in the sequence. Finally, using the estimated periods for the moving objects, we also extract the objects that undergo the corresponding repetitive motions. The segmentation of the jumping girl obtained via spatial correlation is shown in Fig. 9(c).

8 Evaluation Results

We quantitatively measure the performance of our method by estimating the errors in the period estimates and the segmentation (Table 1). The ground truth for the periods of the moving objects is obtained by empirically counting the repetitions of each motion in the sequence. The error e_T in the period estimates T_{est} is then given by the absolute difference of T_{est} and the ground truth T

Table 1. Errors in the Period Estimates and Object Segmentation for 2D Method

Video	e_T (x dir)	e_T (y dir)	e_S for object 1	e_S for object 2
Synthetic	0	0	0.235	0.121
Walking	0	0	0.255	0.27
Swings	0.3	-	0.37	0.443
Jump-rope and Dribbling	0.25	0.4	0.27	0.15

i.e. $e_T = |T_{est} - T|$. When there are many objects in the video, the error in the period estimate in each direction is the mean of their individual errors. The object segmentation ground truth is obtained by manually segmenting out each moving object $S_i(\bar{r})$ and the corresponding error e_S is given by the number of pixels where the extracted and actual objects differ, divided by the number of pixels in the real object area. The segmentation errors are related to the object's real size. They usually originate from blocking artifacts, introduced by the correlation. Since in some experiments there is periodic motion in only one direction or there is only one object, there are some blanks ("−") in the table.

9 Conclusions and Discussion

We have proposed a method for multiple periodic motion estimation that combines frequency and spatial data, to overcome many difficulties and shortcomings of existing purely spatial methods.

1. Our approach detects and estimates multiple periods in a video sequence simultaneously (Sec. 5), in contrast to the existing literature, where each periodic motion is analyzed separately, with the help of manual intervention.
2. The proposed approach can deal with motions that deviate from strict periodicity (Sec. 6), as the mean STFT error is zero. This is also shown in experiments, where the real sequences do not have perfectly periodic motions.
3. Our approach can also extract objects with periodic motion superposed on translation, such as walking (Sec. 5.2). Such motions cannot be analyzed without preprocessing in the existing literature.
4. Once the periods in the video are estimated, the periodically moving objects can be extracted via spatial correlation methods (Sec. 5.1). Since the periods have already been found, our segmentation is more reliable than those of spatial only methods.

References

1. Boyd, J., Little, J.: Motion from transient oscillations. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR 2001. (2001)
2. Seitz, S., Dyer, C.R.: View-invariant analysis of cyclic motion. International Journal of Computer Vision **25** (1997) 231–251
3. Tsai, P., Shah, M., Keiter, K., Kasparis, T.: Cyclic motion detection for motion based recognition. Pattern Recognition **27** (1994) 1591–1603
4. Cutler, R., Davis, L.S.: Robust real-time periodic motion detection, analysis, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 781–796
5. Polana, R., Nelson, R.: Detection and recognition of periodic, nonrigid motion. International Journal of Computer Vision **23** (1997) 261–282
6. Lu, C., Ferrier, N.: Repetitive motion analysis: Segmentation and event classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004) 258 – 263

7. Cohen, L.: Time-frequency distributions-a review. Proceedings of the IEEE **77** (1989) 941–981
8. Pepin, M.P., Clark, M.P.: On the performance of several 2-d harmonic retrieval techniques. In: Signals, Systems and Computers, 1994. 1994 Conference Record of the Twenty-Eighth Asilomar Conference on. (Volume 1.) 254 –258
9. Briassouli, A., Ahuja, N.: Fusion of frequency and spatial domain information for motion analysis. In: ICPR 2004, Proceedings of the 17th International Conference on Pattern Recognition. Volume 2. (2004) 175–178
10. Chen, W., Giannakis, G.B., Nandhakumar, N.: A harmonic retrieval framework for discontinuous motion estimation. IEEE Transactions on Image Processing **7** (1998) 1242–1257
11. Kojima, A., Sakurai, N., Kishigami, J.I.: Motion detection using 3D-FFT spectrum. In: 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing. Volume 5. (1993) 213–216
12. Czerwinski, R., Jones, D.: Adaptive short-time Fourier analysis. IEEE Signal Processing Letters **4** (1997) 42–45
13. Boashash, B.: Estimating and interpreting the instantaneous frequency of a signal - Part 1: Fundamentals. Proceedings of the IEEE **80** (1992) 520–538
14. Djurovic, I., Stankovic, S.: Estimation of time-varying velocities of moving objects by time-frequency representations. IEEE Transactions on Signal Processing **47** (1999) 493–504
15. Kay, S.M.: Modern Spectral Estimation, Theory and Applications. Prentice-Hall, Englewood Cliffs, NJ (1988)