Recognition of Human Periodic Movements From Unstructured Information Using A Motion-based Frequency Domain Approach

Meng, Qinggang; Li, Bo; Holstein, Horst

Recognition of human periodic movements from unstructured information using a motion-based frequency domain approach

Q. Meng [b], B. Li [a,*], H. Holstein [b]

[a] *Department of Computing and Mathematics, Manchester Metropolitan University, Manchester M1 5GD, UK*
[b] *Department of Computer Science, University of Wales, Aberystwyth, UK*

Abstract

Feature-based motion cues play an important role in biological visual perception. We present a motion-based frequency-domain scheme for human periodic motion recognition. As a baseline study of feature based recognition we use unstructured feature-point kinematic data obtained directly from a marker-based optical motion capture (MoCap) system, rather than accommodate bootstrapping from the low-level image processing of feature detection. Motion power spectral analysis is applied to a set of unidentified trajectories of feature points representing whole body kinematics. Feature power vectors are extracted from motion power spectra and mapped to a low dimensionality of feature space as motion templates that offer frequency domain signatures to characterise different periodic motions. Recognition of a new instance of periodic motion against pre-stored motion templates is carried out by seeking best motion power spectral similarity. We test this method through nine examples of human periodic motion using MoCap data. The recognition results demonstrate that feature-based spectral analysis allows classification of periodic motions from low-level, un-structured interpretation without recovering underlying kinematics. Contrasting with common structure-based spatio-temporal approaches, this motion-based frequency-domain method avoids a time-consuming recovery of underlying kinematic structures in visual analysis and largely reduces the parameter domain in the presence of human motion irregularities.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Human periodic motion classification; Motion-based recognition; Gait analysis; Visual perception; Moving light displays (MLDs); Motion power spectral analysis

1. Introduction

Humans communicate large amounts of information via non-verbal body movements and activities, furnishing a rich source of information about intention, emotion and identity. This richness should be exploited when designing user interfaces to intelligent autonomous systems [31]. Visual interpretation of human activity is emerging as an essential and challenging task in machine vision, demanded by potential applications in human–machine interaction, domestic equipment, surveillance systems and the entertainment industry. A large body of research is dedicated to this task using image sequences, as pointed by survey articles [17,29,40]. Remarkable progress in human body detection, tracking, pose estimation and more generally activity recognition, has been

achieved, but usually subject to expensive image analysis of complex articulated movements.

By contrast, the ability of humans to perceive structure and motion from sparse feature point motion cues has been demonstrated by Johansson's Moving light displays (MLDs) [24]. In the MLDs shown in Fig. 1, an image sequence was reduced to a set of moving light dots. These light dots were images of markers, attached at the joint sites of a human subject, contrasted to a dark background. The dots acted as discrete feature-points presenting motion characteristics in the spatio-temporal domain. The MLDs carried only motion information but no structural information, since the displayed points were discrete, unconnected and their identities unknown to the system. One frame of static dots remained meaningless to human observers, while they were able to recognise activities such as walking, running or stair climbing from such sequences conveyed by a small number of lights. Barclay et al. [3] and Cutting and Kozlowski [12] also claimed that human observers can identify an actor's gender and even friends by their gaits in MLDs. These pioneering psychological works relating to human motion perception suggest that

---

* Corresponding author. Tel.: +44 161 247 3598; fax: +44 161 247 1483.
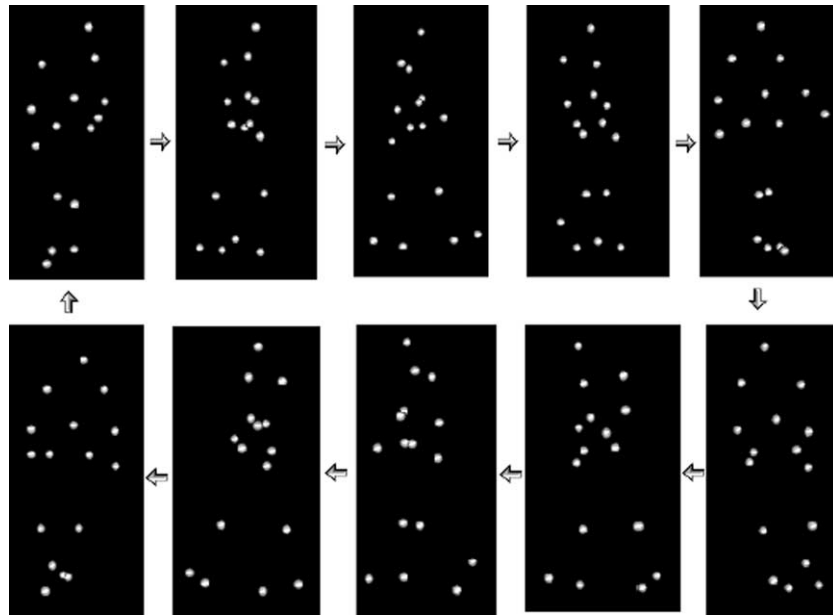*E-mail address:* b.li@mmu.ac.uk (B. Li).

Fig. 1. A clockwise circle-walking person in MLDs with 16 feature points.

feature-based motion cues play an important role in recognition. In the case of machine vision, the biological metaphor suggests that it may be possible to use reduced spatio-temporal information, such as embedded in MLDs, for recognition.

MLDs images, as feature-based motion cues, have been widely used in studies of visual perception [3,12,20,24,28]; human motion tracking and activity recognition in computer vision [6–9,13,19]; clinical gait analysis and sports science research [2,14,37,38,43]; character animation [18,35]; augmented reality and virtual reality [13]. Motion analysis from reduced MLDs allows us to use quantitative, concise and accurate data to investigate essential recognition features in visual perception, motion modelling, kinematic formulation and motion synthesis.

Despite agreement that humans are adept at recognising actions from motion cues in MLDs, there is still no consensus on how humans interpret the MLDs stimuli. Two distinct theories exist. In the first, it is supposed that people use relative motion information in the MLDs and rich pre-knowledge of biological motion and structure to recover the 3D structure of the moving object (person), and subsequently use the structure for recognition—structure-based recognition. In the second, motion information is to be used directly for recognition, denoted by motion-based recognition.

During the last two decades, computer visual interpretation of human activity has emphasised structure-based recognition. Survey Refs. [17,29,40] point to substantial work in this category. Researchers extracted information from images to recover the time varying articulated non-rigid human from. From the recovered underlying structure, high-level motion parameters such as joint angles or trajectories rep resenting the various body part dynamics could then be derived for motion interpretation and recognition. The main problem of structure-based recognition is the high computational cost of explicit articulated structure reconstruction and body part identification, required as a prior necessity. Therefore, structure-based approaches are not readily employed for real-time vision application. To simplify the process of image analysis many studies in the model-based category have employed a subset of body features, identified manually or by using markers, as input for motion interpretation. Campbell and Bobick [17] classified ballet dance steps using a phase space representation. The phase space was related to each degree of freedom derived from identified feature point data on an articulated human body. Goddard [19] proposed a computational model for visual motion recognition of gaits in walking, jogging, and running from MLDs. He used the joint angles and angular velocities as features for recognition. A difficulty in these works was the necessity of prior identified individual points in the images. Goddard has argued the possibility of perception directly from unstructured motion information embedded in MLDs.

Motion-based recognition deals with the recognition of an object and/or its motion, directly from the whole characterised motion pattern in a compact representation, regardless of any underlying structure reconstruction prior to recognition. Shaw and co-workers [9,36] review some representative works in this category. Though relatively few researchers have attempted motion-based recognition, Abdelkader et al. [1] proposed a motion-based structure-free method to characterise motion pattern in monocular video for human gait recognition. Boyd and Little [6], using global shape-of-motion features derived from MLD images, has shown that it is possible to recognise individual people by their gait using non-structural means. Recent work by Wang et al. [41,42], employing spatial–temporal silhouette as biometric motion signature and PAC-based eigenvalue analysis, achieved successful gait recognition from outdoor image sequences in a reduced dimensionality of feature space. These researches avoid the complex vision

problem of kinematic structure recovery and confirm that motion cues play an important role in recognition. We shall summarise motion-based recognition, in particular human periodic motion, in Section 2.

Stemming from cross-fertilisations and achievements drawn from the cross-disciplines of psychology, human visual research and together with the significant development of human motion analysis in computer vision, the central task is to investigate the capability of using feature-based motion cues embedded in MLDs to develop an efficient computational model for human periodic motion recognition, and therefore demonstrate the potential of motion-based recognition by non-structural means. Since the focus of this baseline study is recognition, we do not accommodate bootstrapping from the low-level image processing of feature detection in MLDs images. Instead, we will use a marker-based optical MoCap system to obtain feature-point biological-motion data, which allows us to use the data directly for motion analysis.

We propose in this paper a *motion-based frequency domain* approach for recognition of human periodic movements. The rest of the paper is organised as follows: Section 2 reviews related work on cyclic motion recognition. Section 3 states our method of data collection. Section 4 describes the frequency domain approach for recognition. Section 5 provides experimental results on recognition of human cyclic motion. We discuss and conclude our work in Sections 6 and 7.

2. Human periodic motion recognition

Approaches using motion directly, without regard to its underlying structure, for (human) periodic motion recognition are described in e.g. [1,4,6,10,15,32,33,39,41,42]. Motion-based approaches characterise human periodic motion by, for example, a set of static configurations of the body in each pose in a manner of *state-spaces*, or by analysing shape of motion, trajectories, templates and optical flow images in spatial–temporal dimensions simultaneously. Fourier transforms are often used to detect or recognise periodicity. The detected periodicity is used to assist motion recognition. For such spatio-temporal domain approaches, in order to deal with the problems of human motion irregularities or change in speed, techniques such as *scale space* or *dynamic lime warping* (DTW), considered computationally expensive, are often used for normalisation or for matching portions of scale space to locate similar patterns in state-spaces or templates.

Early work by Polana and Nelson [32] proposed a method of detecting periodic motion using Fourier transforms on several point trajectories. They showed that in principle that the period of the movement could be inferred from averaging the fundamental frequencies of the point trajectories. Tsai et al. [39] used the trajectory of one point of an object performing some cyclic motion to compute trajectory curvature. An autocorrelation was performed to enhance self-similarity within the curvature function. The Fourier transform was finally used to detect the presence of a cycle and its period from the spatio-temporal curvature. Cutler and Davis [11] explored the nature of an object's self-similarity in periodic motion and

applied time–frequency analysis to detect and characterise the periodicity in videos. Fujiyoshi and Lipton [15] generate a 'star' skeleton from the object boundary. They apply Fourier analysis to its skeleton for detecting periodic motion. Then they utilise both posture and motion cycle of the 'star' skeleton to recognise activities such as walking and running. Huang et al. [22] reported a template matching method using eigenspace transformation for feature extraction. An enhanced canonical analysis was employed to reduce feature dimensionality and optimise the separability of different gait classes.

For motion-based methods, motion feature extraction from original images is crucial. Some researchers mark feature points manually or use markers to simplify image analysis for direct recognition [7]. With recent advances in the field of human motion tracking, sophisticated computer vision techniques have been developed to detect joint movement of body kinematics in consecutive frames. Methodologies can be classified into two categories: model-free or articulated-model based. Model-free methods, such as those using 2D contours [5,25,27,34], are usually fast and sometimes real-time, but are compromised in accuracy by their dependence on statistical features related to position, shape, velocity, texture and colour degraded by image noise and body part occlusion. By contrast, articulated-model based methods match an explicit volumetric model to image sequences, particularly in multi-views [16,21,23,30], where motion and stereo measurement of body segments is feasible, accurate and robust. In this sense, many studies, as in the field of gait recognition, combine motion-based recognition with a model-based approach, to assist high fidelity feature detection from images [10,30,44]. For example, Ning et al. [30] employed a simplified human model with enhanced motion constraints for efficient tracking and recognition. Cunado et al. [10] and Yam et al. [44] extracted motion features of the upper leg in video by model-based geometrical matching. Subsequently, a phase-weighted Fourier description was applied to construct frequency domain gait signatures for classification. To counter the inherent recognition difficulty due to human gait irregularity, variability and complexity, they have argued the advantages of encompassing full-body motion signals but noted the difficulty of handling the ever-increasing dimensionality.

The direct application of whole-body frequency domain analysis for motion recognition has received much less attention in video-based analysis. Works emphasising frequency domain analysis are, for example, Angeloni et al. [2] and Köhle and Merki [26]. Angeloni et al. use gait kinematic data from MLDs to analyse the frequency content of whole body movement. Their work presents the characteristic spectral distribution among articulated body parts. Köhle and Merkl demonstrate that the kinetic data from ground reaction force platforms can also be used to classify gait patterns in clinical gait analysis, through Fourier transforms of vertical force components and classification by self-organising maps. The works of both Angeloni and Köhle show that motion frequency spectra may include cues suitable for motion recognition.

We are pursuing a new motion-based method for human motion interpretation. We propose a frequency domain

approach to recognise human periodic motion using unidentified kinematic data from MLDs. As a baseline study, our MLDs data is obtained by a laboratory-based motion capture (MoCap) system, aiming to analyse strategical recognition capabilities rather than the bootstrapping of feature-detection from innate image computing. The central goal was to determine whether or not motion characteristics exist not only in the spatio-temporal domain, but also in the frequency domain; whether or not recognition could be exploited by low-level, non-parametric representations, preserved even in the reduced unstructured MLDs data without recovering underlying geometry by complex image analysis for articulated movements. We demonstrate that the explored motion features in the frequency domain can be effective for classification by non-structural means in the presence of human motion irregularities. The approach is algorithmically and computationally simpler than structure-based spatio-temporal techniques.

3. MLDs data collection

All human kinematic data used in our work are acquired from a marker-based 3D optical motion capture (MoCap) system, the Vicon 512. The system provides 3D coordinates of unidentified trajectories of markers attached to a subject, in the manner of a 3D-MLD system. The data are not affected by the projective distortions of particular camera views. In this respect, we differ from other classical MLD investigations, which detect data from 2D projected image sequences. The available 3D MoCap data allows us to use the data directly for motion analysis without dealing with feature point detection from images in a low level.

In our motion capture system, the world coordinate system has its origin on the ground. The xy-plane is parallel to the ground plane, and the z-axis is vertical. Other conditions for data collection in our experiments are:

- Motions are captured in a control volume, about 4 m (length)$\times$4 m (width)$\times$2.5 m (height). The measurement accuracy is to the order of a millimetre.
- Sixteen markers, regarded as feature points, are attached on human subjects at the following locations (the labels used to indicate them in the following are in the brackets): head (HEAD), anatomical T10 (BACK), shoulders (LSHO, RSHO), elbows (LELB, RELB), wrists (LWRI, RWRI), hips (LASI, RASI), knees (LKNE, RKNE), ankles (LANK, RANK), hallux (LTOE, RTOE). They are effective in indicating motion cues in MLDs.
- The obtained trajectories are nearly always uninterrupted, because the multi-dimensional views furnished by the multi-camera system minimise occlusion events in most motions. Some small trajectory gaps arising from body occlusion are filled by interpolation during MoCap post-processing.
- The correspondence between a 3D trajectory and the marker identity is not assumed known in the motion to be identified. This allows generalisation to harsher scenarios in

which feature point identity is not available. In Figs. 4–7, we have indicated ordered feature point identities purely for display clarity and illustration. Identity information is not used in the recognition process.

4. A frequency domain method

The movements of feature points contain information both of motion and structure identity. For most common periodic activities carried out on a level (horizontal) floor, the vertical components, which are the z-coordinates of 3D-MLD trajectories, imply crucial cues relative to ground ($z=0$) and provide a simple input for Fourier analysis. They can be used without transformation, because they contain no horizontal drift and are motion orientation invariant. In this study, we use only the z-trajectories as the motion cue to be analysed. We find that cues from the unidentified z-trajectories alone suffice to discriminate between a number of simple periodic human activities.

The overall frequency domain schema for modelling periodic movements by feature-point z-trajectories is shown in Fig. 2.

4.1. Power spectral analysis for whole body movement

Our experimental analysis assumes availability only of vertical components of the unidentified trajectories of feature points, $i=1,...,I$, obtained from 3D-MLDs. We apply spectral analysis to the z-component $z_{i(n)}$ of the trajectory of each feature point i of frame samples $n=0,..., N-1$, N being the trial length. The Fourier decomposition of the z-trajectory is expressed by

$$z_{i(n)} = \frac{1}{2}a_{i(0)} + \frac{1}{N}\sum_{k=1}^{N-1} a_{i(k)} \cos(2\pi nk/N) + b_{i(k)} \sin(2\pi nk/N),$$

(1)

where $a_{i(k)}$ and $b_{i(k)}$ are the Fourier coefficients of feature point i in units of millimeter. To achieve an adequate frequency resolution, the length N of each trial is between 256 and 1300 frames, ideally including about 5 gait cycles (G_c) for a specific periodic movement.
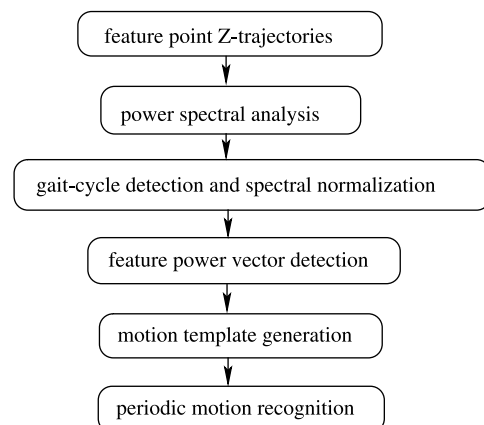

Fig. 2. Modelling periodic movements in frequency domain by z-trajectories.

The power magnitude for the kth frequency harmonic of feature point i is given directly from the Fourier coefficients $a_{i(k)}$ and $b_{i(k)}$ as

$$P_{i(k)} = a_{i(k)}^2 + b_{i(k)}^2, \quad k = 1,2,\ldots,N/2 \tag{2}$$

in units of millimeter square. Examples of such power spectra for a clockwise circle-walking with detected gait cycle (Gc) of 0.97 Hz are given in Fig. 3.

From power spectral analysis of whole-body feature points for a number of common cyclic movements, such as walking, running, jumping, skipping, we find the dominant power of human movements occupies only a narrow bandwidth, with an upper limit of about 10 Hz. The power spectral distribution shows clustering around a fundamental activity frequency and its harmonics [2,10]. The magnitude envelope of a specific power spectrum retains time-shift invariancy regardless of where in time
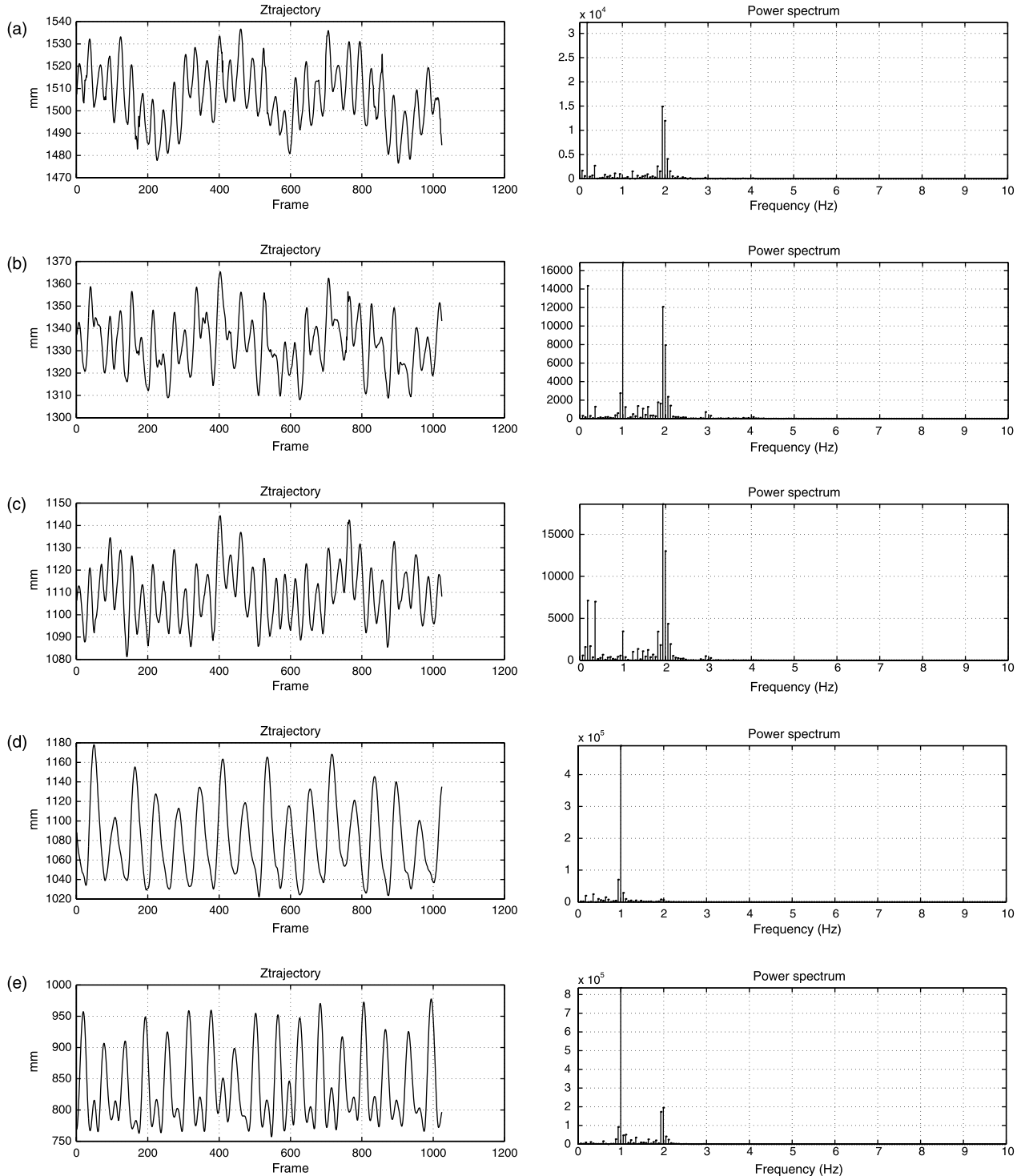
Fig. 3. Examples of vertical-component power spectra of a clockwise circle-walking person. $N = 1024$, $f_\delta = 60$ Hz, $G_c = 0.97$ Hz.
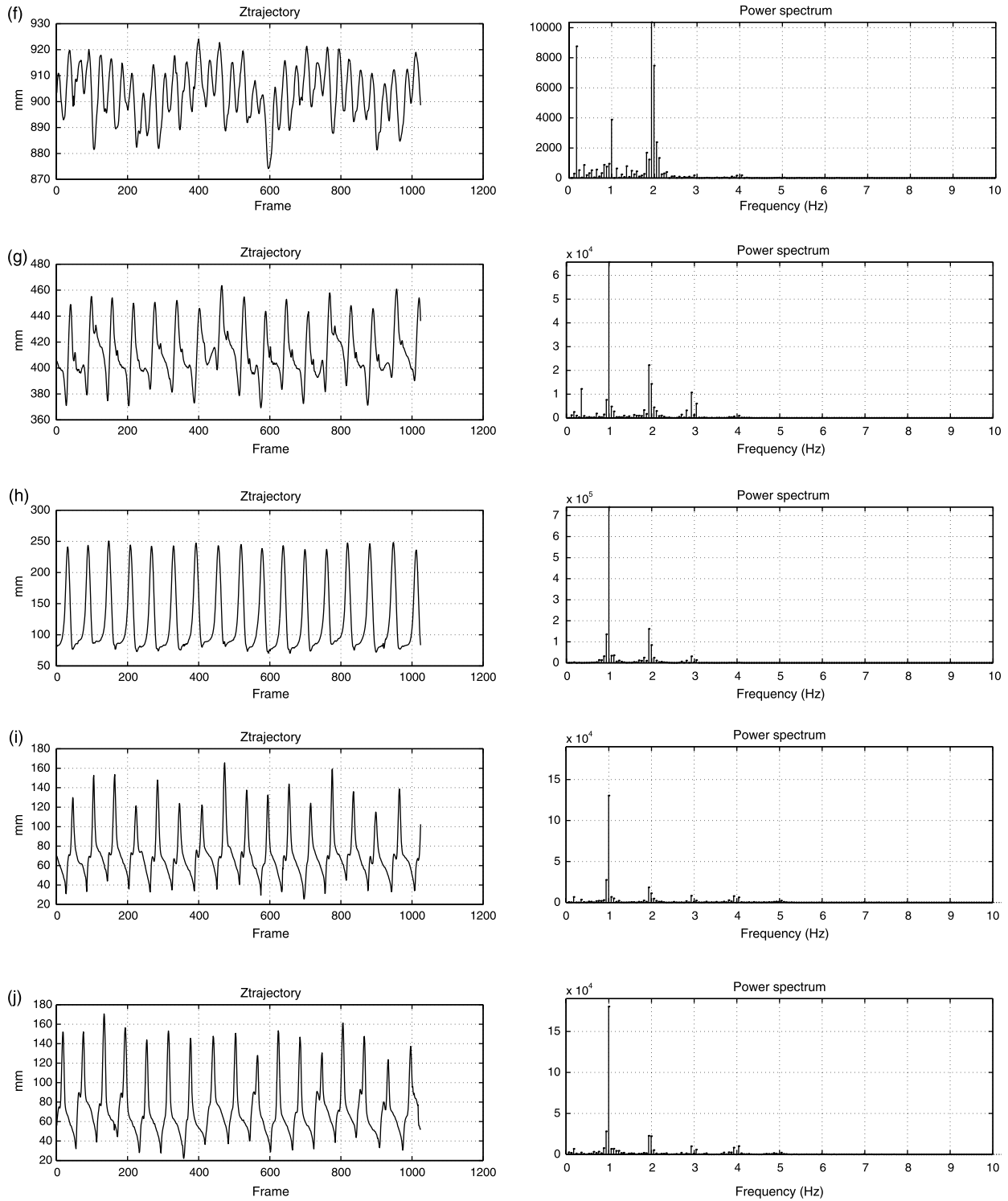
Fig. 3 (*continued*)

the periodic motion is sampled. This characteristic requires no spatio-temporal alignment in frequency domain comparison.

Motion power spectra reflect not only the overall vertical activeness of body parts, consistent with undergoing motion intensity, but also provide a power distribution signature associated with swing/oscillation frequencies underlying the specific motion. For example, as shown in Fig. 3, active body

parts, such as elbow, wrist, knee, ankle and toe, usually exhibit larger power components than relatively steady parts, e.g. head, shoulder, back and hips. The spectra of different feature points present characteristic distributions as well as intra-limb association, e.g. knee–ankle–toe, and elbow–wrist. Full body motion power distribution hints at the possibility of discriminating motion patterns for classification.

Human body parts in skeletal linkage undergoing periodic locomotion present natural rhythmic patterns related to a fundamental activity cycle. This is evident from the power spectra of body parts in Fig. 3. Bi-pedal activities (e.g. walking, running, skipping alterative feet) may exhibit a doubling of the overall activity cycle frequency in certain body parts (e.g. head an hips) [43] with dominant energy around twice the gait cycle. This is observable in the spatio-temporal trajectories of Fig. 3, in which the hip and head cycles appear at twice the knee frequency during walking.

Low frequency components well under the first fundamental in the spectra reflect secular postural changes and human motion irregularity over the trial track. These low frequency noise components are relatively more evident for body parts under-going small vertical movements, such as for head, back and shoulder during walking. The spectra of active body parts, such as elbow, wrist, knee, ankle and toe, show remarkable motion power clustering around the G_c and its harmonics, and a relatively diminished motion noise. We can also observe that the power components of the left (outside) toe (Fig. 3(j)) are larger than that of the right (inside) toe during circle motion Fig. 3(i), though their spectral patterns would show bilateral symmetry in normal forward gait. The power discrepancy arises because the circle walking has larger outside than inside foot movement.

For the same kind of motion in different subjects, spectral patterns of the same feature points are similar, hinting at the motion nature, differences being attributed to variation in individual speed and amplitude. To achieve a speed-invariant representation for the same kind of movement, we normalise whole-body spectra to the fundamental activity cycle or generalised gait-cycle (G_c). To obtain an accurate G_c, we sum corresponding power components over all feature points i at each frequency $k^*\Delta f$ within a band-limited frequency [0.4–5.0] Hz, where $\Delta f = f_\delta/N$ denotes the frequency resolution, f_δ denotes the chosen sample rate, 60 Hz being used in our experiments for human motion. The frequency corresponding to the maximum power magnitude in the first clustering of the resulting spectral sum

$$K^* := \max_k \left\{ \sum_i P_{i(k)} \right\}, \quad k = 1,\ldots,N/2 \tag{3}$$

is regarded as the activity cycle, or generalised gait-cycle ($G_c = K^* \Delta f$).

The detected cycle frequency is subsequently used to normalise the power spectrum frequency axis from Hertz to generalised G_c. Power spectra for different activities with specific speeds are now aligned by fundamental frequency and its harmonics. Fig. 4 shows examples of G_c-scaled whole-body power spectra.[1]

---

[1] In Figs. 4–7, the 16 feature points are arranged in the order of HEAD, BACK, LSHO, RSHO, LELB, RELB, LWRI, RWRI, LASI, RASI, LKNE, RKNE, LANK, RANK, LTOE, RTOE. This particular arrangement is given for illustration purposes only.

4.2. Feature power vector and motion template

Frequency resolution $\Delta f = f_\delta/N$ of a power spectrum will differ for different trial lengths N. There is not a consistent distribution unit of spectral components among these spectra, making impossible a component by component comparison of different trials. A uniform motion template for trials is needed to allow direct comparison of spectral data.

Considering the nature of clustering distributions in power spectra and observing that power magnitudes have insignificant contributions above fourth G_c, we extract a set of dominant power components around G_c and its harmonics to fourth G_c from each spectrum P_i, and regard the result as a *feature power vector* $\vec{v}_i$ of the feature point i:

$$\begin{aligned}
\nu_{i(0)} &= \mathrm{DC}_i = \frac{1}{2} a_{i(0)}, \\
\nu_{i(n)}|_{n=1,\ldots,4} &= \sum_{k \in W_n} P_{i(k)}, \\
\nu_{i(5)} &= \sum_{k \notin W_{1,\ldots,4}, k \neq 0} P_{i(k)}, \\
\nu_{i(6)} &= \sum_{k \neq 0} P_{i(k)}.
\end{aligned} \tag{4}$$

The first element $\nu_{i(0)}$ of the vector $\vec{v}_i$ is the DC component in the Fourier decomposition, denoting the average vertical position of this point. The elements $\nu_{i(n)}$ where $n = 1,\ldots,4$ are representative values reflecting distribution clusters around G_c and its harmonics. To mitigate the frequency resolution problem, we utilise *sum-windows* W_n to sum power components within the range of $\pm 20\%$ G_c around the nth G_c. The $\pm 20\%$ windowing ratio is based on a statistical analysis of a series of spectra. We found 80% of the total power is clustered in this range around G_c harmonics. Moreover, this windowing ratio ensures not only a good separation between G_c harmonics, but also reduces the effect of spectral spreading which might be caused by degraded trajectories due to slightly irregular cycles, inherent data detection noise and possible interpolation (see Section 6). The element $\nu_{i(5)}$ is used to represent the non-selected 'small-power' components that have not been included in the $\pm 20\%$ G_c power windows of $\nu_{i(1)}$ to $\nu_{i(4)}$. The last item $\nu_{i(6)}$ is used to represent the sum of motion powers over all frequencies of feature point i.

DC components and total motion-powers of each feature-point in the example trials of running, skipping and walking are given in Fig. 5. DC components indicate the average vertical positions of feature points relative to the ground-based origin during movements. Large motion powers occur with active limbs, such as wrist, elbow, toe, ankle and knee as exemplified in Fig. 5(b).

To generate a uniform *motion template*, we stack feature power vectors of the I feature points into an $I \times 7$ matrix, $V = \{\vec{v}_i | i = 1,\ldots,I\}$. Each column in the motion template, correspond-ing to a G_c harmonic, is scaled relative to the maximum value in this column. By this means, the power amplitudes are normalised to reduce intra-activity subject dependency. After normalisation, averaging among subjects is used to obtain a single representative
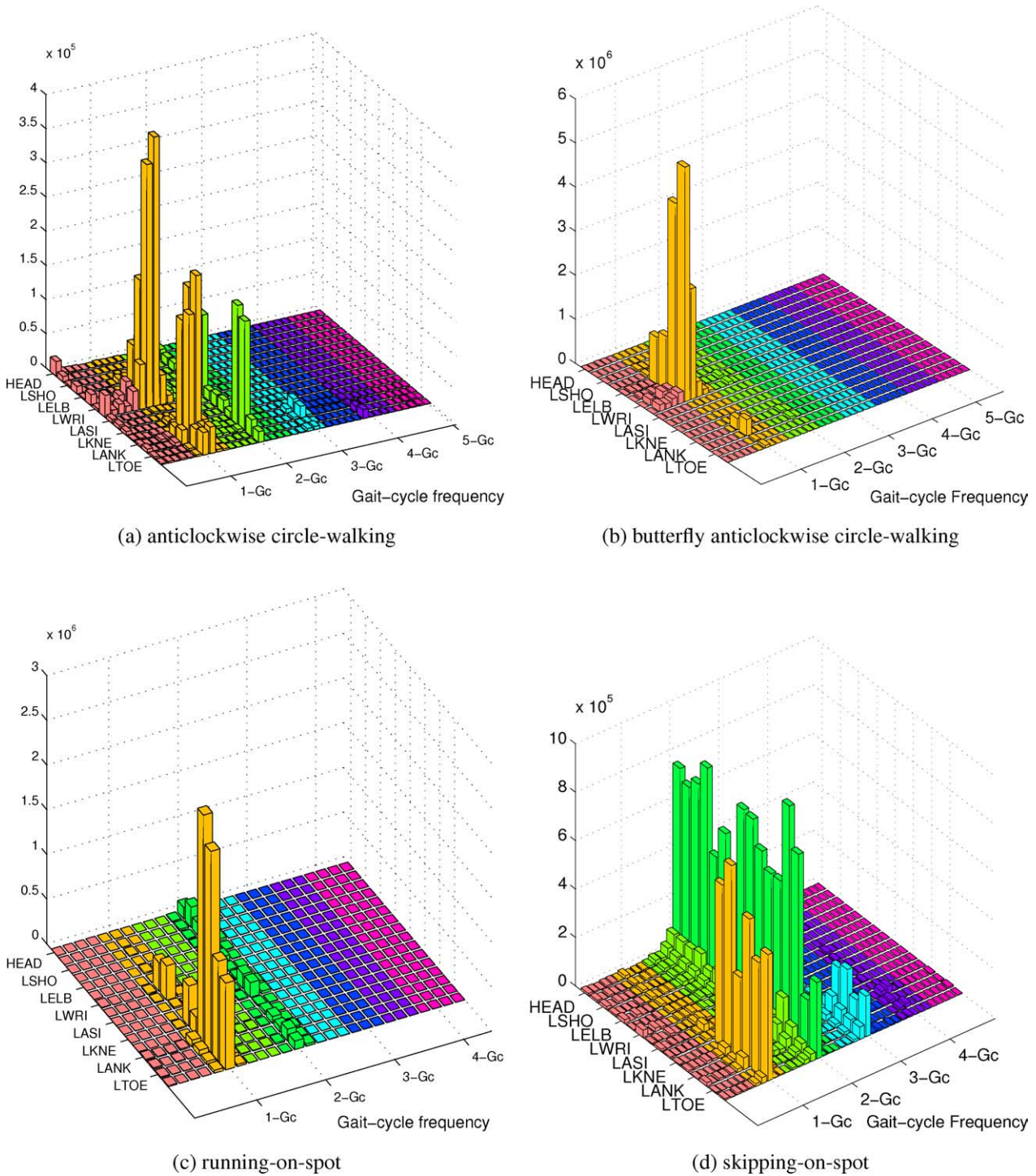
(a) anticlockwise circle-walking

(b) butterfly anticlockwise circle-walking

(c) running-on-spot

(d) skipping-on-spot

Fig. 4. G_c-scaled whole-body power spectra.

standard motion template for a specific motion. Some examples of motion templates are shown in Fig. 6.

Feature power vectors effectively aggregate the frequency components into just seven *feature power elements* ($v_{i(0)}$ to $v_{i(6)}$), a highly reduced dimensionality of feature space. The uniform feature power vector description is now no longer dependent on differently sampled trials associated with specific speeds. This makes direct comparisons of spectral data possible and computationally efficient.

4.3. Motion recognition

Motion recognition is straightforward at this stage. It is achieved by finding the best match between an observed motion template and pre-stored standard motion templates. We apply the algorithm to an observed motion to generate its motion template $U = \{\vec{u}_j | j = 1, \ldots, J\}$, with J unidentified *feature power vectors*. The feature points of the observed motion can be an adequate subset ($J \leq I$) of those used in the
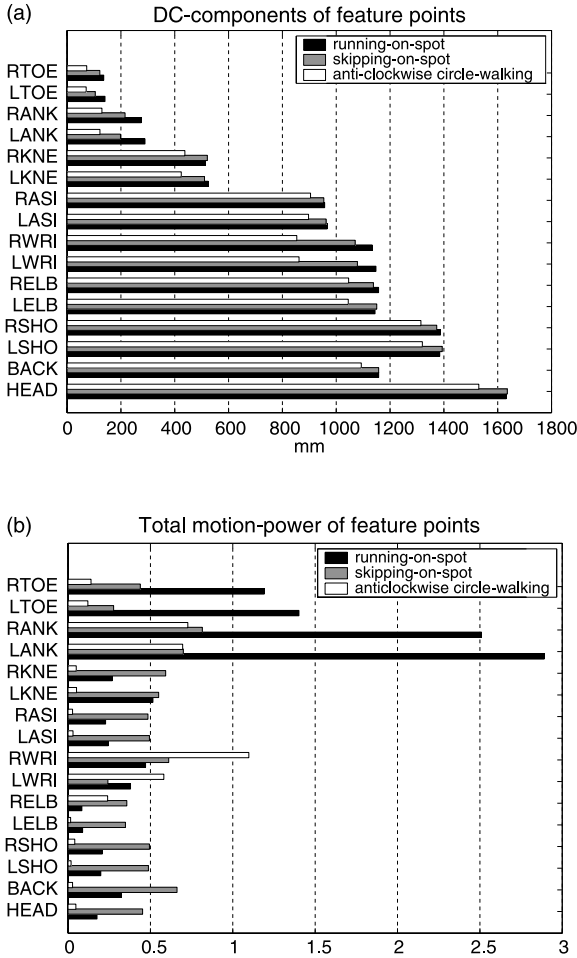
(a)

(b)

Fig. 5. DC components and total motion powers of feature points in three movements.

height information. Human observers had great difficulty in recognising undergoing activities, despite easy identification in standard MLDs. This suggests an appropriate DC weight factor should be assigned along with dominant spectral components. We found best classification results were achieved at a DC weight factor between 0.3 and 0.4 in general.

Relative power distribution ratios of feature power vector components of feature points before normalisation (Eq. (4)) are used to guide weighting factor selection (ω_n, $n = 1,\ldots,5$). Considering all the investigated activities over all trials, we set activity-independent weights to $\vec{\omega} = [0.34, 0.2, 0.2, 0.03, 0, 0, 0.03, 0.2]$ subject to $\sum_{i=0}^{6} \omega_i = 1$. The last parameter ω_6 is used to emphasise the total power intensity which could be weighted equivalently as the dominant components at the first and second G_c. Our weighting ignores the contribution from the fourth G_c ($\omega_4 = 0$), and $\omega_5 = 0.03$ admits a small contribution from the sum of small powers that are omitted by the windowing process around first–third G_c.

The motion match of point j is taken to be the minimum element $\min_i\{m_{j,i}^a\}$ in the jth row of *match matrix* M^a for activity a. This allows motion *power spectral similarity* $\mathcal{S}_P^a$ from all best matches of the J feature points of activity a to be defined by:

$$\mathcal{S}_P^a = \left(1 - \frac{\sum_{j=1}^{J} \min_i\{m_{j,i}^a\}}{J}\right)^3. \tag{6}$$

The motion with maximum similarity $\mathcal{S}_P^a$ for all the searched activity templates is taken to indicate recognition.

Fig. 7 shows two examples of motion recognition indicated by the motion power spectral similarity $\mathcal{S}_P^a$. The observed motion template (with 16 feature points along the vertical axis[2]) is compared with six standard motion templates (with 16 feature points along horizontal axis[2]), respectively, in a manner of *match matrix*. Each 16×16 *match matrix* is intuitively illustrated by a grey-scale graph with 256-levels, in which the progression of white to black denotes increasing difference. For easy observation, we superimposed red on the whitest square to illustrate the best similarity, and green for the second best. For illustration clarity on recognition results, we also arrange feature points of the observed motion in the same order with feature points in the standard templates,[3] but feature-point identity information was not used in the recognition process.

standard templates. We use a $J \times I$ match matrix $M^a = \{m_{j,i}^a | j = 1, \ldots, J, i = 1, \ldots, I\}$

$$m_{j,i}^a = \sum_{n=0}^{6} |\mu_{j(n)} - v_{i(n)}^a| \omega_n, \tag{5}$$

where $\vec{\omega} = [\omega_0, \omega_1, \ldots, \omega_6]$ is a weight vector, to store the weighted difference between each jth motion vector $\vec{\mu}_j$ in the observed template and each ith motion vector $\vec{v}_i^a$ in the model template, for activity a. The weight factors (ω_n, $n = 0, \ldots, 6$) restore the relative importance of the spectral power components before the scaling described in Section 4.2, so that spectral windowed components with larger powers (e.g. around first G_c and second G_c) carry more weight than those of smaller powers (e.g. around third G_c).

We note that the DC component ($n = 0$) in the motion template, indicating the average normalised vertical position relative to the ground-base origin, hints significantly at the pose in motion to infer heuristic structure identity. We demonstrated the importance of the DC component through perception experiments by aligning MLDs of all feature points on the same horizontal reference level, that is, we filtered out the DC

---

[2] Only six labels are displayed, the full list of 16 feature points is given in the footnote [1].

[3] In Fig. 7 and Table 1, motion are named in short as: walk-C for clockwise circle-walking; walk-AC for anticlockwise circle-walking; B-walk-C for butterfly clockwise circle-walking (walking while waving hands up and down); walk-S for walking-on-spot; run-C for clockwise circle-running; run-S for running-on-spot; jump1 for jumping with arms raised to horizontal level, and jump2 for jumping with arms raised over head; skip1 for skipping with feet stepping alternately, and skip2 for skipping with feet stepping together.
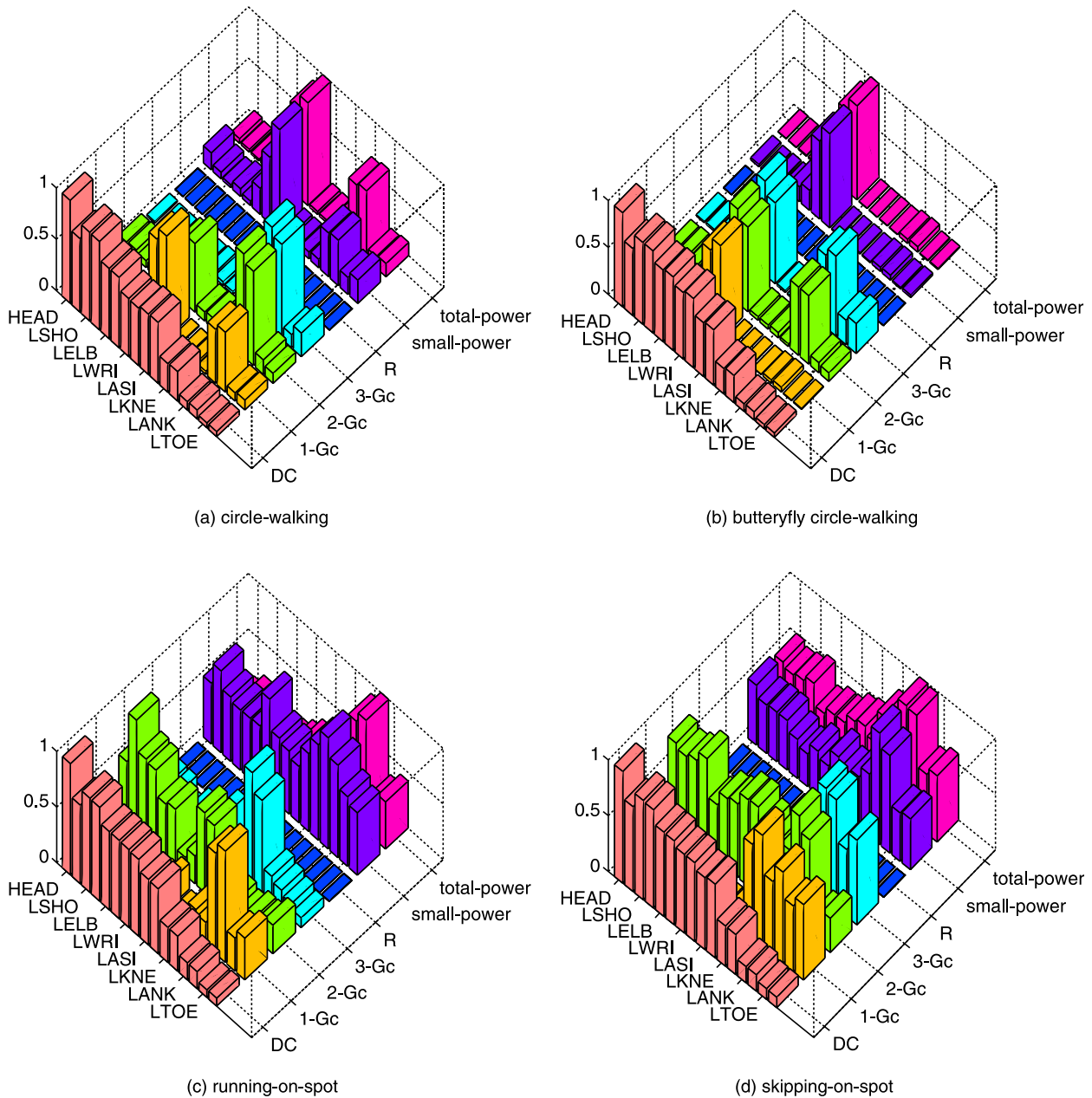
Fig. 6. Examples of motion templates.

In this arrangement, we can easily observe the best S_P^a is reasonably derived from a *match matrix* with a set of red (or green) squares lying along the matrix diagonal. In the correct *match matrix*, green squares usually appear next to red squares, indicating symmetry of corresponding left/right body part movements. In this respect, the *match matrix* may be used to infer not only overall motion similarity, but also imply feature point identity apart from left/right pairings.

5. Experimental results

All experiments were conducted using real motion capture data from a marker-based optical motion capture system, the Vicon 512, as described in Section 3. The motion tracks were captured from a group of 15 subjects that consisted of males and females with ages from 5 to 60 years. Human motion irregularity is inevitable over individuals and trials, though standard activity motion poses were demonstrated to the performers. The trials used for motion template generation were generally separate from test trials used in motion identification.

Recognition was tested on some representative periodic activities, namely walking-on-spot (walk-S), circle-walking (walk-C), clockwise butterfly walking (B-walk-C), running-on-spot (run-S), clockwise circle-running (run-C), skipping type 1 (skip1), skipping type 2 (skip2), jumping type 1 (jump1), jumping type 2 (jump2).
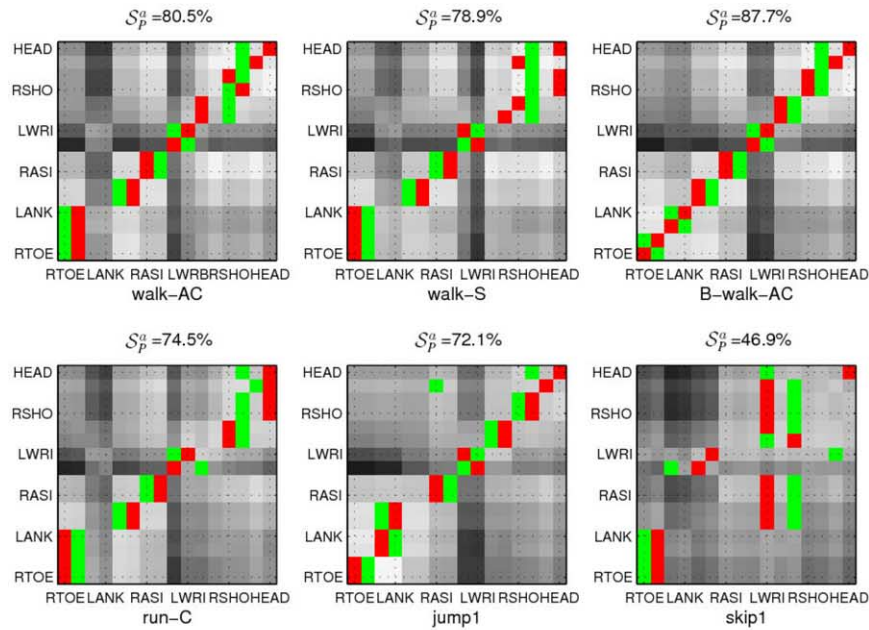
(a) Male subject, butterfly anticlockwise circle-walking.
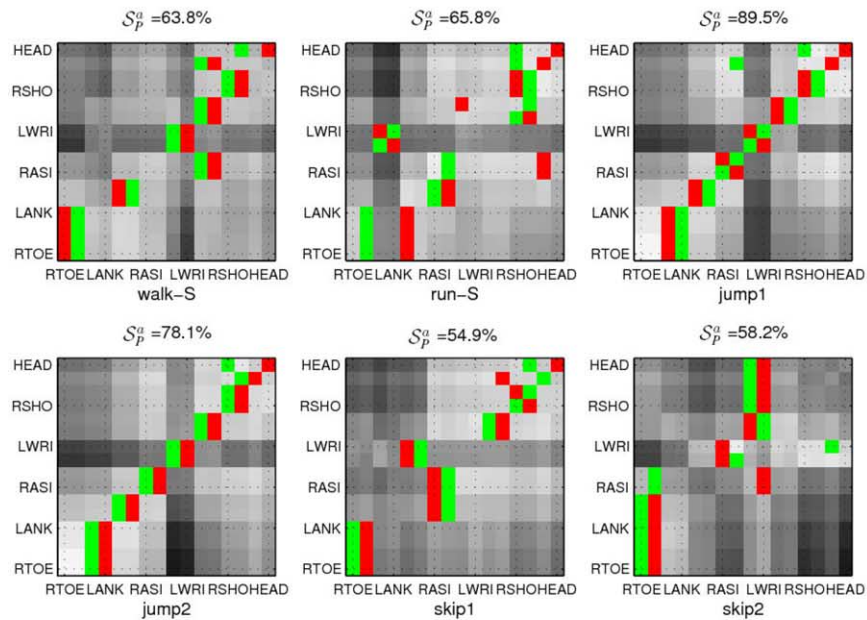
(b) Junior female, jumping-on-spot (jump1).

Fig. 7. Motion recognition indicated by match matrix and $\mathcal{S}_P^a$.

5.1. Recognition by motion power spectral similarity $\mathcal{S}_P^a$

Recognition indicated by *motion power spectral similarity* $\mathcal{S}_P^a$ is shown in Table 1. The averaged similarity parameters in the table matrix indicate the extent of similarity between each type of observed movement (listed in the table column heading 2–10) among each stored motion templates (listed in the leftmost column). The highest column entries, highlighted, indicate the best motion similarity for inferring classification. From the averaged similarity measurement,

we observe the highest column value occurs when the observed activity matches the correct motion template activity. The different classes of movement, such as walking, running, jumping and skipping, are clearly distinguished. Even with similar movements, such as run-spot and run-C, there is discrimination because the magnitudes of power spectra for left and right limbs have a bias in circular activities.

Correct recognition rates for nine types of periodic movements using MoCap data are given in Table 2. We

Table 1
Recognition of human periodic movements by $\mathcal{S}_P^a$ and $\mathcal{S}_{P-Gc}^a$

Activity motion template	Observed activity $\mathcal{S}_P^a(\mathcal{S}_{P-Gc}^a)$								
	Walk-C .75–1.1 Hz	Walk-C .8–1.0 Hz	B-walk-C .8–1.1 Hz	Run-C 1.3–1.5 Hz	Run-S 1.3–1.5 Hz	Jump-1 .9–1.2 Hz	Jump-2 .9–1.1 Hz	Skip-1 .8–1.1 Hz	Skip-2 1.4–1.9 Hz
Walk-C (G_c=0.90 Hz)	**.88 (.87)**	.78 (.82)	.83 (.85)	.68 (.63)	.69 (.63)	.68 (.69)	.70 (.75)	.60 (.65)	.59 (.51)
Walk-AC (G_c=0.90 Hz)	.87 (.86)	.78 (.82)	.83 (.85)	.68 (.63)	.69 (.63)	.67 (.68)	.70 (.75)	.59 (.64)	.59 (.50)
Walk-S (G_c=0.93 Hz)	.81 (.83)	**.79 (.83)**	.80 (.81)	.67 (.64)	.68 (.64)	.66 (.68)	.69 (.75)	.59 (.65)	.58 (.51)
B-walk-C (G_c=0.92 Hz)	.75 (.78)	.72 (.76)	**.90 (.90)**	.60 (.58)	.61 (.57)	.71 (.73)	.72 (.77)	.49 (.56)	.57 (.50)
Run-C (G_c=1.39 Hz)	.73 (.71)	.70 (.69)	.74 (.72)	**.85 (.88)**	.77 (.79)	.70 (.72)	.72 (.72)	.69 (.68)	.58 (.61)
Run-S (G_c=1.42 Hz)	.71 (.69)	.71 (.70)	.63 (.63)	.75 (.78)	**.84 (.86)**	.65 (.67)	.63 (.64)	.70 (.69)	.50 (.53)
Jump-1 (G_c=1.1 Hz)	.58 (.62)	.56 (.62)	.74 (.77)	.63 (.61)	.64 (.63)	**.90 (.91)**	.86 (.86)	.55 (.63)	.53 (.46)
Jump-2 (G_c=0.92 Hz)	.60 (.66)	.62 (.69)	.73 (.76)	.63 (.60)	.63 (.58)	.83 (.82)	**.91 (.92)**	.54 (.62)	.45 (.37)
Skip-1 (G_c=0.97 Hz)	.62 (.64)	.60 (.66)	.48 (.56)	.65 (.63)	.67 (.64)	.52 (.59)	.52 (.60)	**.77 (.81)**	.54 (.46)
Skip-2 (G_c=1.70 Hz)	.63 (.60)	.61 (.60)	.60 (.59)	.61 (.65)	.62 (.67)	.56 (.58)	.54 (.54)	.57 (.56)	**.78 (.80)**

Table 2
Correct recognition rates of periodic movements

Walk-S (%)	Walk-S (%)	B-walk-C (%)	Run-C (%)	Run-S (%)	Jump1 (%)	Jump2 (%)	Skip1 (%)	Skip2 (%)
93	86	94	90	88	95	96	84	85

found best recognition rates occur for jumping, because the designed jumping activities were simple for subjects to execute uniformly, and active body movements enforced frequency domain motion signatures which largely concealed the spectral noise from small posture irregularity. Walking in a circle (walk-C) shows better results than running. This is expected as individual running patterns were more dispersive than walking patterns in our lab-based observations. The B-walk-C gained more credits than walk-C due to the enhanced spectral characteristic of rhythmically exaggerated arm waving in walking. Performers showed motion non-uniformity for some movements open to personal interpretation, such walking or running on spot and especially skipping. Though such subjective factors affected acquisition of ideal periodic motion data in practice, correct recognition has been achieved in most cases.

The experimental results demonstrate that motion comparison using frequency-domain spectral analysis offers effective discrimination among different periodic motions, based solely on unidentified vertical trajectories, in the presence of human motion irregularity.

5.2. Recognition by combined similarity $\mathcal{S}_{P-Gc}^a$

We have found that the parameter $\mathcal{S}_P^a$ reflects motion power characteristics of the whole-body, giving rise to recognition possibility. The parameter $\mathcal{S}_P^a$ has been made insensitive to speed variability for the same activity, by scaling with respect to the G_c. The same scaling, however, has also lost the important discriminating factor of speed among different activities, represented by the value of G_c itself. We therefore considered activity periodicity assisted recognition, defined as *combined motion power spectral similarity* $\mathcal{S}_{P-Gc}^a = \{\mathcal{S}_P^a, \mathcal{S}_{Gc}\}$, formally:

$$\mathcal{S}_{Gc}^a = 1 - \frac{|Gc_{observed}^a - Gc_{model}^a|}{Gc_{model}^a}, \tag{7}$$

$$\mathcal{S}_{P-Gc}^a = 0.8\mathcal{S}_P^a + 0.2\mathcal{S}_{Gc}^a. \tag{8}$$

As shown in Table 1, the combined similarity parameter $\mathcal{S}_{P-Gc}^a$ increases the ability to distinguish motions with substantially different activity periodicity, such as running and walking. It will not help to discriminate motions with similar activity periodicity, such as walking and Jump2 (with activity periodicities G_c around 0.92 Hz).[4]

6. Discussion

The proposed frequency domain approach exhibits a desirable tolerance to human motion irregularity and data measurement noise, because the spectral domain description naturally confines the effect of data corruption to spectral widening while retaining clustered signatures. Such a characteristic has no analogue in spatio-temporal modelling. Retention of special signature can be clearly observed in Figs. 3 and 4. The original MoCap tracks present high irregularity of human motion embedded with MoCap noise, while their spectra show clear clustering patterns hinting at unique motion signatures.

To demonstrate the ability of the proposed spectrum-based recognition approach to handle missing data, possibly a source of serious data degradation, some synthetic experiments were conducted. We randomly cut data from a number of MoCap z-trajectories obtained for different body parts during different kinds of periodic movements. Then, several interpolation methods, such as Linear, cubic spline, cubic Hermite polynomial, were applied to fill the gaps in the corrupted data. Based on the least square comparison between reconstructed and original trajectories, cubic splines provided the best interpolation, and were employed to investigate the

---

[4] Model template G_cs are given in the leftmost column, and the range of observed G_cs are shown in the subsequent column headings, Table 1.

effect of track reconstruction on their spectral decomposition. We found small gaps can be accurately filled by cubic splines to give spectral fidelity and do not compromise recognition.

In Fig. 8, we show some examples in synthetic situation. Original MoCap hip and knee z-tracks for walking-on-spot include about 20 gait cycles. For each track, 30 gaps were randomly generated with average gap size of 25% gait period. The simulated total missing data is therefore about 38% of the trajectory length. Cubic spline interpolated trajectory segments are shown by red dotted lines in the figure. We observe the reconstructed spectra, even for the highly distorted hip track shown in Fig. 8(c), maintain broad fidelity to the original. This demonstrates the robustness of spectral analysis for handling trajectory distortion, and the potential accuracy of motion template matching.

From Fig. 8, we note that spectral distortion arising from degraded track data by interpolation over larger gaps is most likely to appear as power spread around G_c harmonics and added small-power terms which could ambiguiate the boundary between G_c harmonics. This is because band-limited interpolation essentially works as a low-pass filter. The windowing method used for detecting feature motion vectors has been designed to reduce such spectral ambiguity (see Section 4.2). Meanwhile, for motion template comparison and recognition, the small-power term is less weighted in order to reduce the influence of spectral noise arising particularly from
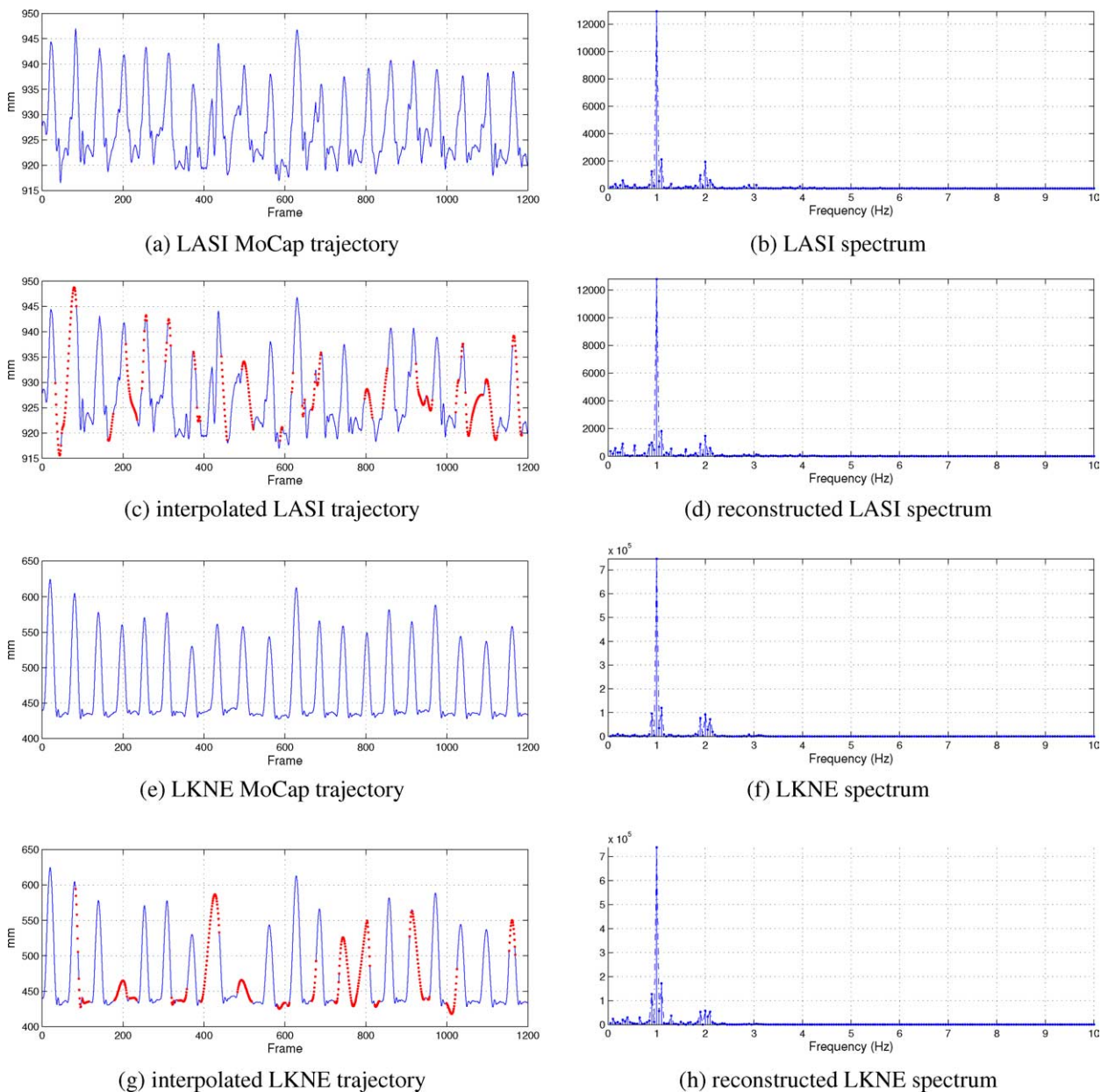

(a) LASI MoCap trajectory

(b) LASI spectrum

(c) interpolated LASI trajectory

(d) reconstructed LASI spectrum

(e) LKNE MoCap trajectory

(f) LKNE spectrum

(g) interpolated LKNE trajectory

(h) reconstructed LKNE spectrum

Fig. 8. Effect of simulated data distortion though interpolation on power spectra in a case of walking-on-spot, $G_c = 1$ Hz. (Cubic spline interpolated trajectory segments are shown by red dotted lines.)

the small-power signals of trajectory noise (see Section 4.3). From the synthetic data distortion experiments, we found that recognition based on the proposed frequency-domain motion cues not unduly affected by interpolation.

As a feasibility study at this stage, we utilised an activity-independent weight vector for motion template matching. We noted from our experiments that activity-dependent weight vector could be used to improve recognition. This is left for the future work.

7. Conclusions

We proposed a motion-based frequency domain approach based on the spectral analysis of whole-body motion data sampled at selected feature points to discriminate and recognise human periodic activities. The approach demonstrates the feasibility of feature-based motion cues for recognition by utilising unidentified kinematic data from MLDs. Full-body power spectral analysis applied to the vertical-trajectory components was found to be adequate to furnish motion cues, obviating the need for costly horizontal movement analysis. *Feature power vectors* are detected to efficiently code a *motion template*, as an activity signature averaged for a number of subjects, for indexing each kind of motion. Recognition is carried out by motion template comparison of an observed motion with standard motion templates to find a best match.

In addition, the frequency domain approach has by nature a robustness to spatio-temporal corrupted data arising form human motion irregularity and measurement noise through spectral clustering. Heuristic methods were investigated to exploit this frequency domain attribute. Feature power vectors separately aggregate trivial and massive raw spectral components into a small number of numerical measures that effectively retain clustering signatures and confine the effect of power spreading. The uniform description makes direct comparisons of spectral data possible in a condensed parameter dimension. Normalisation both on frequency and power magnitude allows template matching to be independent on differently sampled trials associated with specific speeds and be carried out for a wide range of subjects. The choice of feature points is not a priori prescribed. The only requirement is that the chosen feature points effectively reflect motion cues and are common to all templates, and that the observed movement is based on all or a subset of the feature points used in template construction.

We have found that inherent characteristics of human periodic movements exist in the algorithmically simple yet computationally efficient frequency domain, contrasting previous work in the spatio-temporal domain. Frequency domain features hint at motion nature in a manageable parameter domain in the presence of human motion irregularities, allowing effective classification from low-level, reduced information embedded in MLDs by non-structural means.

The experiences gained from the present study of using MLDs data suggest research extension to more complex activity modelling including individual recognition by motion-based methods. Much of this baseline study using concise MLDs data would be transferable to the harsher scenario of image sequences, subject to feature data acquisition which could take advantage of recent advance in human motion tracking and image analysis technique. This research is consistent with, and could contribute to, the important research areas of biologically inspired machine vision.

References

[1] C. Abdelkader, R. Cutler, L. Davis, Motion-based recognition of people in EigenGait space, in: IEEE International Conference on Automatic Face and Gesture Recognition, 2002.
[2] C. Angeloni, P.O. Riley, D.E. Krebs, Frequency content of whole body gait kinematic data., IEEE Transactions on Rehabilitation Engineering 2 (1) (1994) 40–46.
[3] C.D. Barclay, J.E. Cutting, L.T. Kozlowski, Temporal and spatial factors in gait perception that influence gender recognition, Perception and Psychophysics 23 (2) (1978) 145–152.
[4] A. Bobick, J. Davis, The recognition of human movement using temporal templates, IEEE Transactions on PAMI 23 (3) (2001) 257–267.
[5] A. Bobick, J. Davis, The recognition of human movement using temporal templates, IEEE Transactions on PAMI 23 (3) (2001) 257–267.
[6] J. Boyd, J. Little, Global versus structured interpretation of motion: moving light displays, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 1997.
[7] L. Campbell, A. Bobick, Recognition of human body motion using phase space constraints, in: Proceedings of the IEEE International Conference Computer Vision, Cambridge, 1996, pp. 624–630.
[8] C. Cédras, M. Shah, A survey of motion analysis from moving light displays, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, Washington, June 1994, pp. 214–221.
[9] C. Cédras, M. Shah, Motion-based recognition: a survey, Image and Vision Computing 13 (2) (1995) 129–155.
[10] D. Cunado, M. Nixon, J. Carter, Automatic extraction and description of human gait models for recognition purposes, Computer Vision and Image Understanding 90 (1) (2003) 1–41.
[11] R. Cutler, L. Davis, Robust periodic motion and motion symmetry detection, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, South Carolina, USA, June 2000, pp. 326–331.
[12] J.E. Cutting, L.T. Kozlowski, Recognizing friends by their walk: Gait perception without familiarity cues, Bulletin of the Psychonomic Society 9 (5) (1977) 353–356.
[13] K. Dorfmüller, Robust tracking for augmented reality using retroreflective markers, Computers and Graphics 23 (1999) 795–800.
[14] G. Ferrigno, M. Gussoni, A procedure to automatically classify markers in biomechanical analysis of whole body movement in different sport activities, Medical and Biological Engineering and Computing 26 (1988) 321–324.
[15] H. Fujiyoshi, A. Lipton, Real-time human motion analysis by image skeletonization, in: Proceedings of the IEEE Workshop on Applications of Computer Vision, 1998.
[16] D. Gavrila, L. Davis, Model-based tracking of humans in action: a multiview approach, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, San Francisco, 1996, pp. 73–80.
[17] D.M. Gavrila, The visual analysis of human movement: a survey, Computer Vision and Image Understanding 73 (1) (1999) 82–98.

[18] M. Gleicher, Animation from observation: motion capture and motion editing, Computer Graphics 33 (4) (1999) 51–55.

[19] N.H. Goddard, The perception of articulated motion: recognizing moving light displays, PhD Thesis, University of Rochester, 1992.

[20] H. Hill, F. Pollick, Exaggerating temporal differences enhances recognition of individual from point light displays, Psychological Science 11 (2000) 223–228.

[21] A. Hilton, D. Beresford, T. Gentils, R. Smith, W. Sun, J. Illingworth, Whole-body modelling of people from multi-view images to populate virtual worlds, International Journal of Computer Graphics 16 (7) (2000) 411–436.

[22] P. Huang, C. Harris, M. Nixon, Recognising humans by gait via parametric canonical space, Journal of Artificial Intelligence in Engineering 13 (4) (1999) 359–366.

[23] A.D.J. Deutscher, I. Reid, Automatic partitioning of high dimensional search spaces associated with articulated body motion capture, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2001.

[24] G. Johansson, Visual motion perception, Scientific American 232 (6) (1975) 75–80 (see also pages 85–88).

[25] I.A. Kakadiaris, D. Imitri Metaxas, 3D human body model acquisition from multiple views, International Journal of Computer Vision 30 (3) (1998) 191–218.

[26] M. Köhle, D. Merkl, Things we observed when watching people walk: classification of gait patterns with self-organizing maps, in: Proceedings of the Seventh Australian Conference on Neural Networks ACNN'96, Canberra, April 1996.

[27] Y. Li, A. Hilton, J. Illingworth, A relaxation algorithm for real-time multiview 3D-tracking, Image and Vision Computing 20 (12) (2002) 841–859.

[28] Y. Ma, H. Paterson. A. Dolia, S. Cho, A. Ude, F. Pollick, Toward a biologically-inspired representation of human affect, Brain Inspired Cognitive Systems, 2004.

[29] T.B. Moeslund, E. Granum, A survey of computer vision-based human motion capture, Computer Vision and Image Understanding 81 (3) (2001) 231–268.

[30] H. Ning, L. Wang, W. Hu, T. Tan, Articulated model based people tracking using motion models, in: Proceedings of the IEEE International Conference on Multimodal Interfaces, 2002.

[31] A. Pentland, Smart rooms, Scientific American 274 (4) (1996) 68–76.

[32] R. Polana, R. Nelson, Detecting activities, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 1993, pp. 2–7.

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

[34] X. Ren, A. Berg, J. Malik, Recovering human body configurations using pairwise constraints between parts, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2005, pp. 824–831.

[35] J. Richards, The measurement of human motion: a comparison of commercially available systems, Human Movement Science 18 (5) (1999) 589–602.

[36] M. Shah, R. Jain, Motion-based Recognition, Kluwer, Dordecht, 1997.

[37] I. Söderkvist, P.A. Wedin, Determining the movements of the skeleton using well-configured markers, Journal of Biomechanics 26 (12) (1993) 1473–1477.

[38] A.J. Stoddart, P. Mrázek, D. Ewins, D. Hynd, Marker based motion capture in biomedical application, IEE Electronics and Communications 103 (1999).

[39] P.S. Tsai, M. Shah, K. Keiter, T. Kasparis, Cyclic motion detection for motion based recognition, Pattern Recognition 27 (12) (1994) 1591–1603.

[40] L. Wang, W. Hu, T. Tan, Recent developments in human motion analysis, Pattern Recognition 36 (3) (2003) 585–601.

[41] L. Wang, T. Tan, W. Hu, H. Ning, Automatic gait recognition based on statistical shape analysis, IEEE Transactions on Image Processing 12 (9) (2003) 1120–1131.

[42] L. Wang, T. Tan, H. Ning, W. Hu., Silhoutte analysis based gait recognition for human identification, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (12) (2003) 1505–1518.

[43] M. Whittle, Gait analysis: An Introduction, Butterworth-Heinemann, Oxford, 1996.

[44] C. Yam, M. Nixon, J. Carter, Automated person recognition by walking and running via model-based approaches, Pattern Recognition 37 (5) (2004) 1057–1072.