

On Parameterizing Higher-order Motion for Behaviour Recognition

Yan Sun^{a,b,*}, Jonathon S. Hare^c, Mark S. Nixon^c

^a*Computer Engineering and Science School, Shanghai University, 99 Shangda Road, Shanghai, 200444, PR China*

^b*Shanghai Institute for Advanced Communication and Data Science, 333 Nanchen Road, Shanghai 200444, P.R. China*

^c*Electronic and Computer Science, University of Southampton, University Road, Southampton, SO17 1BJ, United Kingdom*

Abstract

Human behaviours consist different types of motion; we show how they can be disambiguated into their components in a richer way than that currently possible. Studies on optical flow have concentrated on motion alone without the higher order components: snap, jerk and acceleration. We are the first to show how the acceleration, jerk, snap and their constituent parts can be obtained from image sequences, and can be deployed for analysis, especially of behaviour. We demonstrate the estimation of acceleration in sport, human motion, traffic and in scenes of violent behaviour to demonstrate the wide potential for application of analysis of acceleration. Determining higher order components is suited to the analysis of scenes which contain them: higher order motion is innate to scenes containing acts of violent behaviour, but it is not just for behaviour which contains quickly changing movement: human gait contains acceleration though approaches have yet to consider radial and tangential acceleration, since they concentrate on motion alone. The analysis of synthetic and real-world images illustrates the ability of higher order motion to discriminate different objects under different motion. Then the new approaches are applied in heel strike detection in the analysis of human gait. These results demonstrate that the new approach is ready for developing new applications in behaviour recognition and provides a new basis for future research and applications of higher-order motion analysis.

Keywords: Motion Analysis, Higher-order Motion, Acceleration, Jerk, Snap

** Corresponding author.*

Email addresses: yansun@shu.edu.cn (Yan Sun), jsh2@ecs.soton.ac.uk (Jonathon S. Hare), msn@ecs.soton.ac.uk (Mark S. Nixon)

1. Introduction

Human behaviours consist of many different types of motion. In the simplest sense, walking can be easily distinguished from running by acceleration and many behaviours have more complicated motion. There has been much sophisticated work on activities recognition based on optical flow. Chaudhry et al. proposed histograms of oriented optical flow for action recognition [2]. Kumar and John utilize optical flow vectors on the action edges to recognize human actions and interactions [3]. Kolekar and Dash fuse shape and optical flow together for activity recognition [4]. Wang et al. match feature points between frames using SURF descriptors and dense optical flow as their hand craft features [59]. More recently, Edison and Jiji have proposed two optical acceleration algorithms and utilize them to recognize action but do not extend beyond acceleration [58]. In this paper we extend higher-order motion analysis and parameterizes different types of motion for behaviour recognition based on our previous work [5], [6]. The new contributions of this paper are:

- Optical flow field is disambiguated systematically into acceleration, jerk and snap fields and their components. Our acceleration algorithm achieves superior performance to the state of art on efficiency and preciseness.
- We show generalised application of higher-order motion detection in a richer selection of imagery for not only the new approach but also the new basis, on synthetic and on real image sequences.
- Jerk and snap flow are applied for detecting heel strikes and the results demonstrates that jerk can be adapted in gait analysis or other applications on real images.
- We demonstrate the capability to discriminate violence using velocity and acceleration flow and show the advantages of using analysis of acceleration.

This paper is arranged as follows: Section 2 gives a brief review of background gait analysis and optical flow. Section 3 describes the higher-order motion algorithms. Section 4 evaluates the algorithms on a rich selection of imagery for behaviour and then for heel strike detection in human gait. Section 5 gives the preliminary results on using acceleration components for violence detection. Discussion and future work for this new computer vision approach are in Section 6.

2. Related Work

2.1. Activity Recognition

Activity recognition has been subject to increasing research interest and extensively studied in recent years. The main applications are automated surveillance analysis including gait analysis [7] and anomaly behaviour detection [8], ambient-assisted living such as gesture recognition [9] and medical assistant [10]. The techniques can be classified by sensor-based and video-based. Sensor-based techniques obtain the information from accelerometer, gyroscope while video-based methods analysis the data from images. Liu et al. utilize the temporal relationships between actions to classify activities from sensor-based data [11]. Dollar et al. develop sparse spatio-temporal features for detecting and characterizing behaviour from video sequence [60]. Ma et al. propose a twin stream network to integrate hand appearance, object attribute and local hand motion to recognize activities [12]. Du et al. utilize metric learning to locate potential anomaly target in a scene [13].

Gait analysis is one of the main branches of activity recognition. In biometric gait can be obtained at a distance and hard to hide or disguise compare with face, fingerprint or iris [14]–[16]. There has been some work achieved the purpose on detecting acceleration data by accelerometers and gyroscopes in gait analysis. Djurić-Jovičić et al. [17] use accelerometers to measure the angle

of leg and ankle, Rueterbories et al. [10] capture the angular displacement to detect gait events via gyroscopes. Yam et al. proposed an analytical gait model which extracts the angle of the thigh and lower leg rotation without parameter selection [18]. Świtoński extracted the velocity and acceleration of skeleton root element, feet, hands and head from gait path as the gait feature [19]. These study have been utilized physical-based acceleration information for gait analysis, optical flow-based acceleration is yet to be investigated.

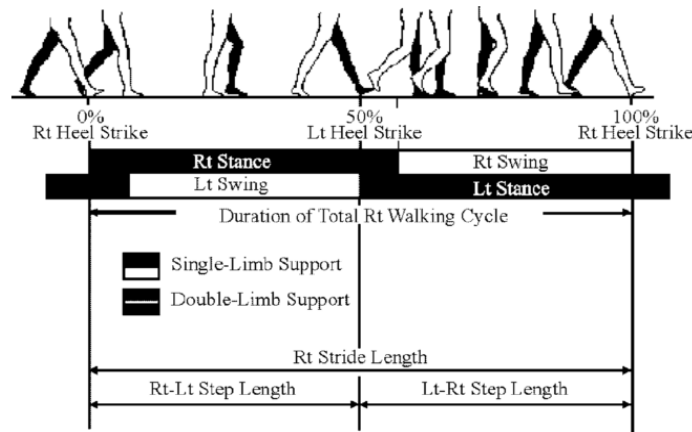


Fig. 1. The temporal components contained in a gait cycle and step and stride length during the cycle [20].

Gait is periodic and most analytic approaches rely on accurate gait phase detection. The components of one gait cycle are shown in Fig. 1: a gait cycle is defined as the interval between two consecutive heel strikes of the same foot therefore heel strike is one of the most important components which partition walking sequences into cycles composed of stance and swing phases [21]. A heel strike refers to the moment the heel first strikes the floor. Suppose one gait cycle starts from the heel strike of right foot, the right foot rotates on the heel to touch the floor (‘stance phase’) to support the body while the left foot is swinging forward (‘swing phase’) until the left heel strikes the floor. Then the roles of the two feet switch, the left foot remains flat on the floor whilst the right foot is swinging forward. When the right heel strikes the floor again, then a gait cycle is complete. Bouchrika and Nixon [22] extracted low-level features, corners, to estimate the strike

positions. Jung and Nixon use the movement of the head to detect the key frame (the frame in which the heel strike takes place) [23].

2.2. *Optical Flow Estimation*

Optical flow estimation is one of the earliest and still active research topics in Computer Vision. Many methodological concepts have been introduced and the performances have been improved gradually since Horn and Schunck proposed the first variational optical flow estimation algorithm, however, the basic assumption has changed little [1]. Most state-of-art approaches estimate optical flow by optimising the weighted sum of two terms: a data term and a prior term [24].

The most popular data term is brightness constancy, it constrains the intensity of one point on the image pattern to be constant between successive frames. The alternative forms of the data term are the correlation between frames [25] and colour space [26]. Apart from the intensity, robust pairwise features also can be used to construct the motion fields. Brox et al. combine gradient constancy with brightness [27]; DeepFlow [28] and SIFT flow [29] both use the Scale-Invariant Feature Transform (SIFT) as the pairing feature.

For prior term smoothness is the most popular choice which assumes that the neighbouring pixels have a similar motion. Trobin et al. use a second-order prior to constrain the smoothness to achieve high-accuracy optical flow [30]. Wedel et al. adapt rigid motion as their prior [31]. Naturally, learning algorithms have started to feature in recent years. Sun et al. [32] learned a statistical model of both the error of brightness and smoothness constraints. FlowNet [33] use Convolutional Neural Networks (CNNs) to predict optical flow from a large quantity of training data. In the next section, we will review previous study of optical flow which has aimed to analyse acceleration.

2.3. Previous Higher-order Motion Algorithms

There was little work analysing higher-order motion before our work on determining gait events through acceleration flow [34]. Chen et al. [35] established an acceleration detection algorithm (without an implementation) based on Lucas-Kanade [36]: the brightness is constant in three consecutive frames and acceleration flow is smooth over a small patch, which turns the constraints into an over-determined equation. In the only other contemporaneous approach, Dong et al. [37] obtain acceleration fields by computing optical flow on optical flow yet it is difficult to obtain spatial partial derivatives in a smoothed area. Edison and Jiji proposed Horn Schunck Optical Acceleration (HSOA) [57] by applying brightness and smoothness constraints. Later, they optimized the algorithm by expanding each frame as a quadratic polynomial, namely Farneback optical flow (FBOA) [58]. However, the constraints are too strong once it extends to multiple frames in a highly dynamic environment.

3. Disambiguating Optical Flow into Higher-order Components

3.1. Estimating Acceleration Flow

New algorithms for optical flow estimation continue to emerge, and the performance has significantly improved since the first variational algorithm by Horn and Schunck [38]. Rather than obtaining flow from its basis we use a state-of-art algorithm, DeepFlow [28], as our basis approach to approximate the acceleration flow. This applies the brightness constancy assumption in small deformable patches, which improves the performance of large displacement estimation and non-rigid matching significantly. The basic approach can be replaced with any new optical flow algorithms, and could thus improve the performance of our higher-order motion algorithms in the

future. The acceleration field $\hat{\mathbf{A}} \in \mathbb{R}^2$ can be estimated by the difference of the neighbouring velocity fields:

$$\hat{\mathbf{A}}(t) = \frac{d\mathbf{V}(t)}{dt} = \mathbf{V}(t, t + \Delta t) - \mathbf{V}(t - \Delta t, t) \quad (1)$$

where $\mathbf{V}(t, t + \Delta t) \in \mathbb{R}^2$ denotes velocity field, it contains the horizontal and vertical components between t and $t + \Delta t$. $\mathbf{V}(t - \Delta t, t)$ is the velocity field between $t - \Delta t$ and t . In the implementation, optical flow field is considered as the velocity field due to the fixed frame rate, with units of pixels/frame.

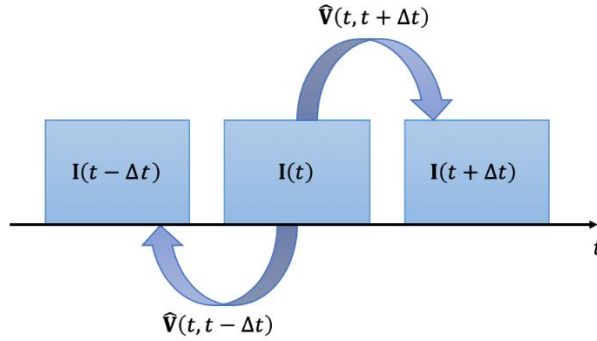


Fig. 2. Computing acceleration field by referring to the middle frame as the start.

To avoid the noise caused by inconsistent referencing in time axis, the start frame in other word, Differential-Acceleration refers to the middle frame as the initial time in temporal template as explained in Fig. 2 [34], the acceleration field $\hat{\mathbf{A}}(t)$ is approximated by:

$$\hat{\mathbf{A}}(t) = \mathbf{V}(t, t + \Delta t) - (-\mathbf{V}(t, t - \Delta t)) = \mathbf{V}(t, t + \Delta t) + \mathbf{V}(t, t - \Delta t) \quad (2)$$

where $\mathbf{V}(t, t + \Delta t)$ denotes the velocity field estimated by DeepFlow [28] from frame t to $t + \Delta t$, and $\mathbf{V}(t, t - \Delta t)$ is the velocity field from t to $t - \Delta t$.

3.2. Beyond Acceleration

It is well known that velocity measures the change in position over time, and acceleration is

the change of velocity, *jerk* describes the change of acceleration [39]. Acceleration links a force acting on a mass from Newton's Second Law:

$$\vec{F} = m\vec{a} \quad (3)$$

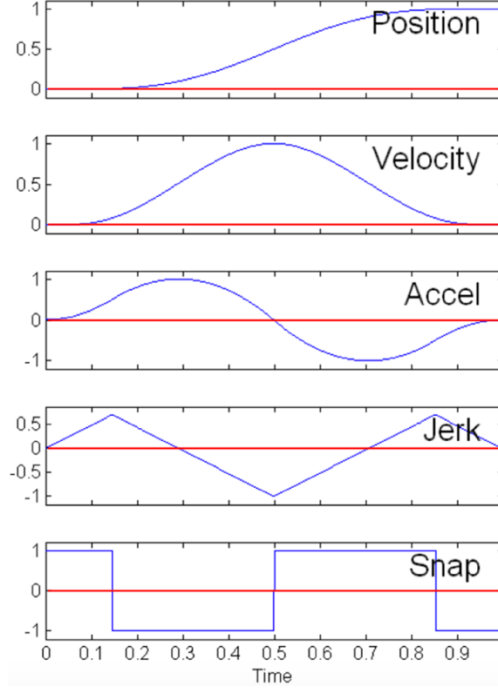


Fig. 3. The relationship between motion profiles in a straight linear motion [45].

Hence in kinematics, assuming constant mass, *jerk* describes the change of force; and *snap* describes the change of jerk. In calculus, snap is the derivative of acceleration with respect to time, the fourth order derivative of position [39]. The relationship between the motion components is given in Equation (4), based on the time evolution of position \vec{r} :

$$\vec{s}(t) = \frac{d\vec{j}(t)}{dt} = \frac{d^2\vec{a}(t)}{dt^2} = \frac{d^3\vec{v}(t)}{dt^3} = \frac{d^4\vec{r}(t)}{dt^4} \quad (4)$$

where \vec{s} represents snap, \vec{j} denotes jerk, and \vec{a} , \vec{v} , \vec{r} , t are acceleration, velocity, position and time respectively. The change of n^{th} -order flow under limited snap is shown in Fig. 3 illustrating the relationship between them.

The conventional application of jerk and snap is in motion control since humans have limited tolerance to change of force. Motion limitation is therefore needed to avoid users losing control during transportation; 2.0 m/s^{-3} in a straight-line transportation is an acceptable limit for most people. The most common examples are in elevator and vehicle design. Now acceleration and jerk have been widely used to analyse the driving behaviours in intelligent driving evaluation: predicting potential risks and guarantee the passengers' comfort in the autonomous driving system [40], [41]. Bagdadi and Varhelyi found that the breaking jerk of vehicles measured by accelerometer is highly related to accidents, their evaluation system based on jerk is 1.6 times better than the longitudinal acceleration methods [42].

In road and track design, unbounded radial jerk needs to be avoided on curved parts: the theoretical optimum strategy is to linearly increase the radial acceleration. Another application using acceleration and jerk is the evaluation of operational path of numerical control machines [43]. In 2006, Caligiuri et al. used jerk to monitor how drug-induced side effect affect patients' handwriting [44]. More recently, a detection algorithm for manoeuvring targets using radar has considered analysing jerk [46, 47].

Most previous approaches derive jerk from the trajectories of moving objects. Datta et al. [47] computed acceleration and jerk vectors from the trajectory of objects' head for person-on-person violence detection. Zaki et al. [48] determine vehicles' jerk from the trajectory extracted by Kanade-Lucas-Tomasi Feature Tracker algorithm [49]. Following our Differential-Acceleration, jerk field $\hat{\mathbf{J}}(t)$ is computed by differencing the neighbouring acceleration fields:

$$\hat{\mathbf{J}}(t) = \mathbf{A}(t, t + \Delta t) - \mathbf{A}(t - \Delta t, t) \quad (5)$$

where $\mathbf{A}(t, t + \Delta t)$ and $\mathbf{A}(t - \Delta t, t)$ denote the acceleration flow fields from frame t to $t + \Delta t$ and $t - \Delta t$ to t respectively by Eq. (2).

Differential-Acceleration refers to the middle frame to avoid the inconsistent start positions since it requires three points to obtain acceleration flow. Estimating snap requires five points which makes referring the middle frame as the start possible (in a similar manner with acceleration flow). The snap field $\hat{\mathbf{S}}(t)$ therefore is computed in a format similar to Differential-Acceleration:

$$\hat{\mathbf{S}}(t) = \mathbf{J}(t, t + \Delta t) + \mathbf{J}(t, t - \Delta t) \quad (6)$$

where $\mathbf{J}(t, t + \Delta t)$ and $\mathbf{J}(t, t - \Delta t)$ are the estimated jerk fields between frame t to $t + \Delta t$ and t to $t - \Delta t$.

3.3. Decomposing the Resultant Flow into Radial and Tangential Components

Presenting acceleration, jerk and snap using their radial and tangential components can help understanding. Acceleration flow was decomposed in our previous work [5]. Further, jerk and snap flow are now resolved as well: the definition of tangential and radial components of jerk are the linear changes of tangential and radial components of acceleration, they are computed in a manner similar to Equation (5) and (6):

$$\begin{aligned} \hat{\mathbf{J}}_{tan}(t) &= \mathbf{A}_{tan}(t, t + \Delta t) - \mathbf{A}_{tan}(t - \Delta t, t) \\ \hat{\mathbf{J}}_{rad}(t) &= \mathbf{A}_{rad}(t, t + \Delta t) - \mathbf{A}_{rad}(t - \Delta t, t) \end{aligned} \quad (7)$$

where $\hat{\mathbf{J}}_{tan}$ and $\hat{\mathbf{J}}_{rad}$ represent the estimated tangential and radial jerk fields, \mathbf{A}_{tan} and \mathbf{A}_{rad} are the tangential and radial acceleration fields. Snap fields are decomposed in a manner similar to Equation (7). Now we have the algorithms to determine the change of acceleration and jerk, and their tangential and radial components. In the next section these shall applied to synthetic and real images to determine whether they can be used to reveal different motion features reliably.

4. Experiments

4.1. Demonstrating Acceleration and Higher-order of Motion Flow on Synthetic Image Sequence

A set of synthetic images was created to simulate the motion of Newton's cradle (a device consisting of a set of swinging spheres, it is used for demonstrating conservation of momentum and energy). The simple motion of Newton's cradle is a good for demonstrating the difference between various flow fields due to the different order in which components change over time.

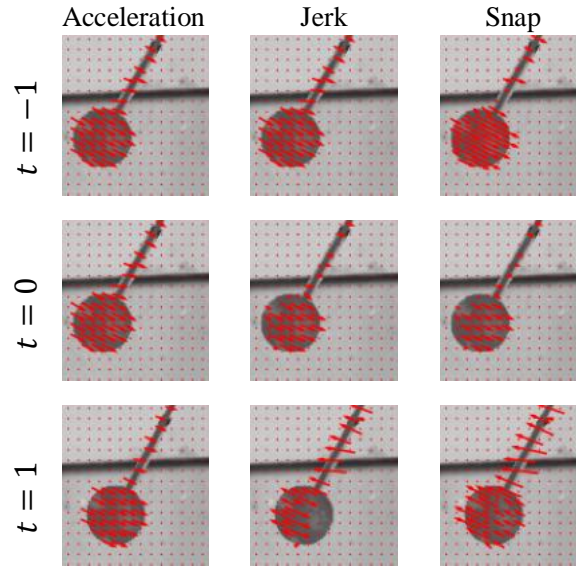


Fig. 4. Flow fields of synthetic cradle sequence.

The results of n^{th} order motion flow of $t = -1, 0, 1$ with particular detail are presented in Fig. 4. When $t = -1$, the sphere is swinging to the left until reaches the highest position (deaccelerating) and remains there for one frame ($\mathbf{v} = 0, t = 0$), then it swings back at $t = 1$ (accelerating). The acceleration flow shows similar character in all three frames where the sphere is accelerating to the lower right with a similar magnitude. On the other hand, jerk and snap flow give different perspectives. When $t = -1$, jerk and snap flow point to the same direction with acceleration and the magnitudes of snap are much larger which reveals that the acceleration is increasing. Both jerk and snap have noticeable reduction when $t = 1$. In the last row the cradle is

swinging back to the centre. The directions of jerk and snap are opposite to the acceleration proves that now the acceleration is decreasing.

4.2. Results on the Middlebury Benchmark

Table 1 shows a performance comparison with Dong’s acceleration algorithm [37], HSOA [57], FBOA [58] and ours (Differential-Acceleration). It reports the Average End-point Error (AEPE) between the estimated flow vector and the pseudo ground truth. The pseudo ground truth is obtained by MDP-Flow2 [50] since the ground truth on the evaluation website only contain optical flow field between two consecutive frames for preventing the upcoming new algorithms to overfit the test images. AEPE is chosen here to evaluate the performance as the ground truth motion in Middlebury image sequences is relatively large. Chen’s algorithm [35] is not included in the results due to the lack of implementations for the other algorithms. Table 2 reports the mean runtime (100 runs) of each algorithm on a single CPU for 640×480 images, our acceleration algorithm is twice faster than Dong’s [37]. It also includes the time cost of Jerk and Snap flow.

Table 1 AEPE of estimation algorithms.

Algorithm	Dong [37]	HSOA [57]	FBOA [58]	Differential-Acceleration
AEPE (pixels)	3.2	4.58	0.67	0.34

Table 2 The runtime (mean of 100 runs) of each algorithm for 640×480 images.

Method	Runtime (seconds)
Dong [37]	37.618
Differential-Acceleration	16.001
Jerk	16.003
Snap	16.008

Some details of the results are shown in Fig. 5. In the first column Beanbags, the optical flow field contains all the types of motion so they appear chaotic lacking any sensitivity to different types of motion. In contrast the tangential and radial acceleration fields indicate the change of the direction and magnitude of each component. The tangential jerk field has more dense flow than radial jerk on the left arm which reveals that the main change of acceleration is magnitude rather than direction. In Basketball, the left man is passing the ball to another person. The ball has not started spinning so the acceleration is mainly tangential. Although the radial acceleration is sparse, the majority directions of the flow on the ball and its shadow on the wall point to the centre of the trajectory. The direction of radial jerk is opposite to the radial acceleration indicates that the spinning acceleration of the basketball is decreasing. The results suggest that higher-order motion flow can help distinguish objects undergoing different types of motions. Again, there is a lot of flow although sometimes it is hard to interpret due to the complexity of real motion.

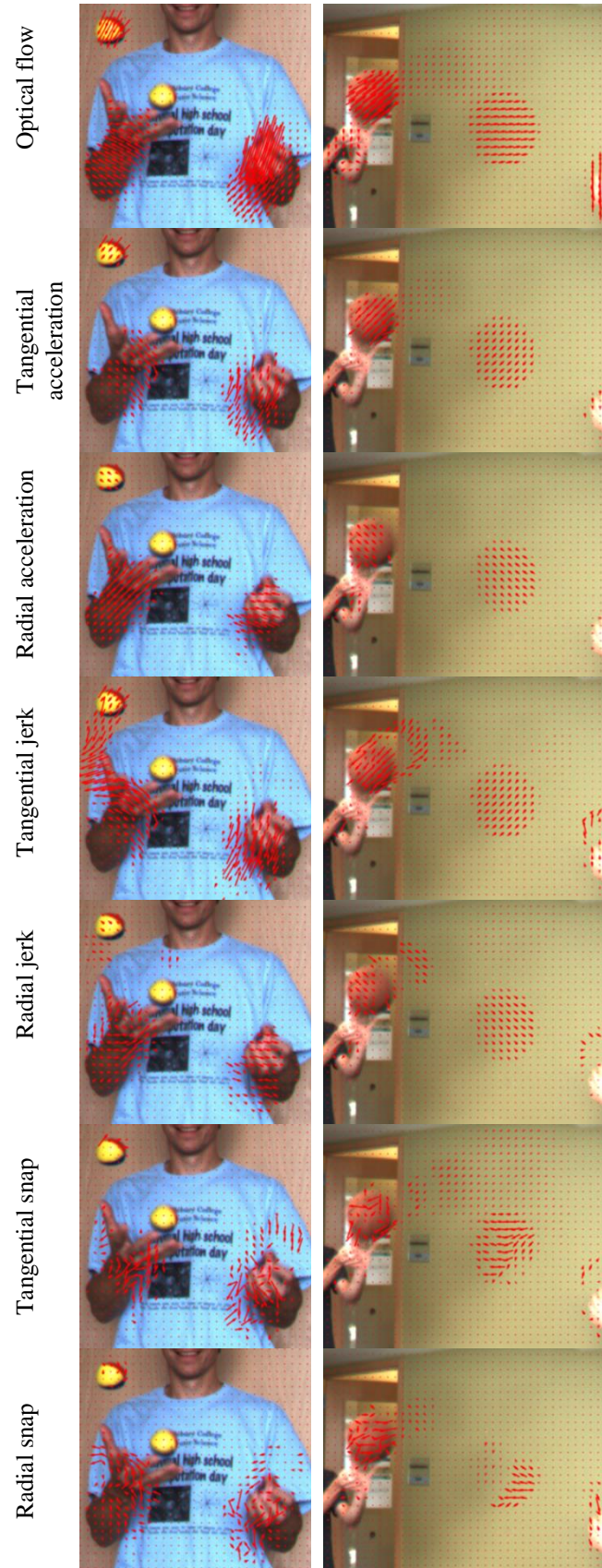


Fig. 5. Discriminatory capability of different estimates of acceleration vs optical flow.

Dumptruck gives another encouraging example in Fig. 6. The silver car and the red dump truck are both moving in constant velocity (approximately), and the other two cars are both accelerating after passed the intersection. In contrast optical flow in (a) cannot distinguish between the accelerating objects, it detects all the moving cars whereas acceleration flow is able to recognize the accelerating objects (the cars in front of the red truck). The results demonstrate that our algorithm can differentiate the accelerating objects from those which are simply moving. The results demonstrate the power and ability of our Higher-order Motion Flow decomposition algorithms. In the next section, they are applied to detect heel strikes for gait analysis.

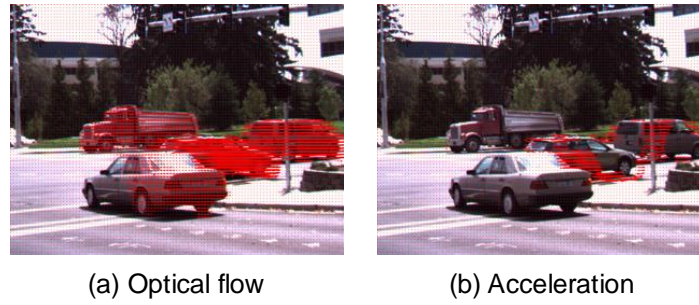


Fig. 6. The flow fields of **Dumptruck**.

4.3. *Heel Strike Detection via Multi-order Flow*

During walking the human torso moves like several connected pendula and researchers have successfully simulated pathological gait by using a liner inverted pendulum model [52,53]. A pendulum has a regular acceleration pattern, which implies that gait can be described by the acceleration pattern of the video-based data. In our previous work [6], the key frame of heel strike can be determined by the quantity of acceleration flow within the ROI. The strike positions can be found from the centres of rotation caused by radial acceleration. In this work, we show normalized

radial jerk flow field of a walking cycle sampled every 7 frames by our new higher-order motion flow algorithms in Fig. 7. There is a considerable amount of flow on the leading foot periodically when the heel strikes on the floor although snap is relatively noisy. Logically, we wonder whether jerk and snap are able to detect and localize the heel strikes or improve the performance than acceleration.



(a) Jerk



(b) Snap

Fig. 7. Patterns of higher order acceleration exist within a gait cycle.

Algorithm 1. Heel strike detection system via radial jerk flow.

for frame in video:

$\text{acc_1} = \text{acc_algorithm}(\text{frame_1}, \text{frame_2}, \text{frame_3})$

$\text{acc_2} = \text{acc_algorithm}(\text{frame_2}, \text{frame_3}, \text{frame_4})$

$\text{rad_acc_1}, \text{centre_map_1} = \text{decomp_components}(\text{acc_1})$

$\text{rad_acc_2}, \text{centre_map_2} = \text{decomp_components}(\text{acc_2})$


```

rad_jerk = rad_acc_2 - rad_acc_1
centre_map = k1*centre_map_1 + k2*centre_map_2

ROI = extract_region(silhouette)
for each_pixel in ROI:
    if rad_jerk[each_pixel] > magnitude_thres:
        rad_amount += 1
    else:
        pass
end for

while rad_amount is peak:
    "KEY FRAME!"
    strike_position = desity_accumulation(centre_map[ROI])
end while
end for
return key_frames_num, strike_positions

```

The framework of our heel strike detection system is given in Algorithm 1. In the pseudo code, the radial jerk field is derived from radial acceleration field based on Equation (5), radial snap flow following the similar manner with Equation (6). The detection system are evaluated on the Large Gait Database (SOTON) [53] and The OU-ISIR Gait Database (OU-ISIR) [54], [55]. SOTON was built in 2002 by the University of Southampton, Shutler et al. [53] collected walking sequences from over 100 subjects, including indoor (controlled lighting) and outdoor (uncontrolled lighting). Since the database focused on the key factors that affect gait recognition other than background segmentation, Chroma-keying was applied in indoor data recording, which is used in the experiments. OU-ISIR is another benchmark dataset for gait recognition, it upgraded the scale of subjects in a gait database significantly. Currently it is the largest gait database in the world: it contains 10,307 subjects (5,114 males and 5,193 females) in total. Their ages range from 4 to 89 years [55], which is two times larger than their earlier version in 2012 [54]. More than 100 heel strikes in each scenario and the test data incorporates multiple viewing angles and walking directions with gait sequences recorded indoors and outdoors.

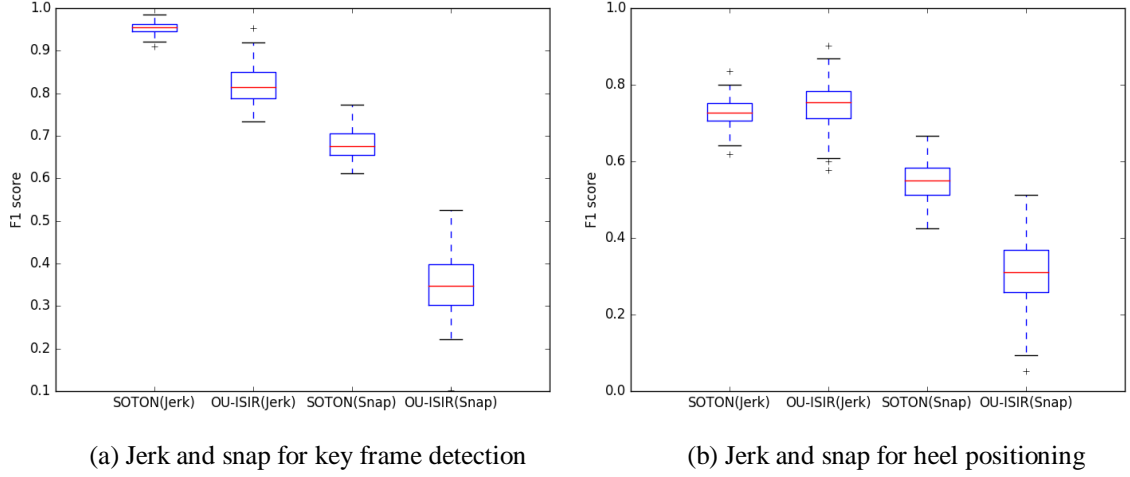


Fig. 8. F1 score of heel strike detection via jerk and snap.

Fig. 8 reports the F1 score of detecting heel strike through jerk and snap. The criterion for a true positive key frame detection is if the detected frames are within ± 2 frames from the manually labelled ground truth. For heel positioning a distance within ± 10 pixels (along both axes) from the ground truth is considered as a true positive. Jerk shows competitive ability on key frame detection and slightly lower on positioning. This indicates that jerk can be adapted for gait analysis or other applications on real images. On the other hand, as a higher order of motion snap underperforms than jerk, also the relatively small ROI in OU-ISIR increases the detection difficulty for jerk. The PR curves in Fig. 9 suggest that jerk has a high area under the curve, which indicates it hits both a high recall and a high precision. It performs considerably better than snap on the balance between precision and recall rate.

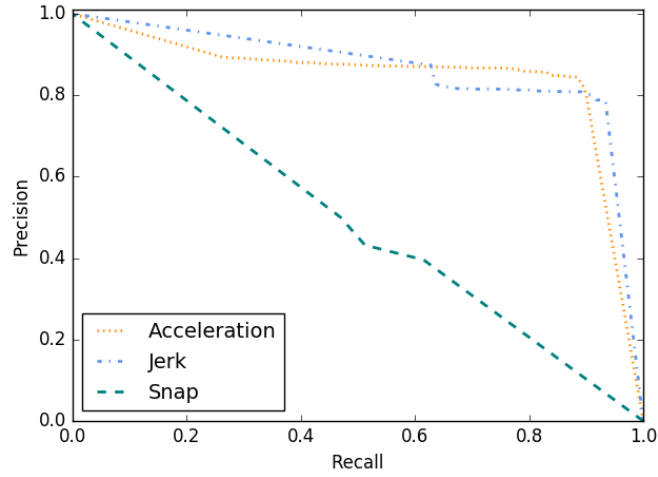


Fig. 9. Key frame detection PR curves of Jerk and Snap.

5. Detecting Violent Behaviour

When people fight, their bodies tend to have large acceleration (in many places and with large magnitudes) on their bodies because their arms are swinging and their feet are kicking. As such, acceleration appears more suited to the detection of rapid change, consistent with scenes of violence. Chen et al. [35] apply acceleration flow to detect abnormal behaviours, Dong et al. [37] feed optical flow and acceleration flow into a network, both in combination and individually, to detect the violence in scenes. They illustrate that the most reliable feature is individual acceleration flow.





Calm scenes

Scenes of violence

Fig. 10. The difference between velocity and acceleration within scenes of calm and violence.

We show the acceleration fields with the optical flow on some raw videos from YouTube and Hockey Fight Dataset [56]. There is no fighting and the scenes are mundane in the left two columns in Fig. 10. We only present single images here, though these are derived from image sequences. In the prison surveillance video (the first row), there is little acceleration detected revealing only the swinging arm of a uniformed guard in the mundane episode. In comparison there is more optical flow, consistent with more leisurely movement as prisoners receive their visitors from the left. This is motion consistent with a tranquil scene; velocity is much smoother than acceleration of which there is none. In the ice hockey video, optical flow shows on the background rather than the skating athlete since the camera is tracking him and the acceleration field is clear. The third

row shows the early stage of an assault before violence, the fourth row is a subway scene before an assault and the last row concerns a robbery. In all these examples velocity exceeds the acceleration and in most cases the acceleration is little.

The right two columns in Fig. 10 incorporate various scenes of violence which are the other parts of the image sequences shown in the first two columns. Optical flow fields are illustrated for contrast. In the episodes of violence there is considerable detected acceleration and much less focus on irrelevant subjects. The first row shows an assault in a prison scene and there is considerably more acceleration than velocity and the acceleration appear to be consistent with kicking and punching. In the second row, acceleration only focus on the people who involve fighting (the two athletes rather than the referee), whereas the velocity flow cannot distinguish them. The assault in the third row shows the velocity and the acceleration appear to be of similar magnitude, but the acceleration is concentrated around the limbs. In the last episode the criminal flees after the crime, their body also tends to make more acceleration. Thus, by detecting acceleration we might be able to determine an approach suited to the detection of violent crime in the future.

6. Conclusions and Future Work

An image sequence compounds many different orders of motion. Considering the diversity of motion, this paper systematically classifies motion fields into different levels for computer vision. As the frame sampling rate in video sequence is fixed the motion fields between neighbouring frames is usually considered as velocity. Acceleration flow was derived from the basis of Horn-Schunck in our earlier work [5] and now it has been extended into the detection of jerk and snap (and in their vector format). The new algorithms are investigated both on synthetic and real images,

providing a new angle to understand and disambiguate dynamic relationships in image sequences. The synthetic cradle demonstrates the difference between higher-order motion characters at its critical status. Furthermore, the experiments of test image sequences from Middlebury dataset and heel strike detection show that the new extensions have the power to further discriminate higher orders of acceleration successfully. The nature of higher order motion detection suggests that the techniques might be more susceptible to noise, as this can be exacerbated when detecting higher order motion. Clearly the new approaches are ripe for further evaluation and application, perhaps in video inpainting, activity recognition or video classification, or in more general image sequence analysis.

Reference:

- [1] D. Fortun, P. Bouthemy, and C. Kervrann, "Optical Flow Modeling and Computation: A Survey," *Comput. Vis. Image Underst.*, vol. 134, pp. 1–21, 2015.
- [2] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1932–1939.
- [3] S. S. Kumar and M. John, "Human activity recognition using optical flow based feature set," in *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, 2016, pp. 1–5.
- [4] M. H. Kolekar and D. P. Dash, "Hidden Markov Model based human activity recognition using shape and optical flow based features," in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2017, pp. 393–397.
- [5] Y. Sun, J. S. Hare, and M. S. Nixon, "Analysing Acceleration for Motion Analysis," in *2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 2017, pp. 289–295.
- [6] Y. Sun, J. S. Hare, and M. S. Nixon, "Detecting heel strikes for gait analysis through acceleration flow," *IET Comput. Vis.*, vol. 12, no. 5, pp. 686–692, Aug. 2018.
- [7] Y. Zhang, G. Pan, K. Jia, M. Lu, Y. Wang, and Z. Wu, "Accelerometer-Based Gait Recognition by Sparse Representation of Signature Points with Clusters," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1864–1875, 2015.
- [8] D. Singh and C. Krishna Mohan, "Graph formulation of video activities for abnormal activity recognition," *Pattern Recognit.*, vol. 65, pp. 265–272, May 2017.
- [9] J. C. Núñez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Vélez, "Convolutional

- Neural Networks and Long Short-Term Memory for Skeleton-based Human Activity and Hand Gesture Recognition,” *Pattern Recognit.*, vol. 76, pp. 80–94, Apr. 2018.
- [10] J. Rueterbories, E. G. Spaich, B. Larsen, and O. K. Andersen, “Methods for gait event detection and analysis in ambulatory systems,” *Med. Eng. Phys.*, vol. 32, no. 6, pp. 545–552, 2010.
 - [11] Y. Liu, L. Nie, L. Liu, and D. S. Rosenblum, “From action to activity: Sensor-based activity recognition,” *Neurocomputing*, vol. 181, pp. 108–115, 2016.
 - [12] M. Ma, H. Fan, and K. M. Kitani, “Going Deeper into First-Person Activity Recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1894–1903.
 - [13] B. Du and L. Zhang, “A discriminative metric learning based anomaly detection method,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6844–6857, 2014.
 - [14] G. Orrù, G. L. Marcialis, and F. Roli, “A novel classification-selection approach for the self updating of template-based face recognition systems,” *Pattern Recognit.*, vol. 100, pp. 107–121, Apr. 2020.
 - [15] W. J. Wong and S.-H. Lai, “Multi-task CNN for restoring corrupted fingerprint images,” *Pattern Recognit.*, vol. 101, pp. 107–203, May 2020.
 - [16] K. Nguyen, C. Fookes, R. Jillela, S. Sridharan, and A. Ross, “Long range iris recognition: A survey,” *Pattern Recognit.*, vol. 72, pp. 123–143, Dec. 2017.
 - [17] M. D. Djurić-Jovičić, N. S. Jovičić, and D. B. Popović, “Kinematics of gait: New method for angle estimation based on accelerometers,” *Sensors*, vol. 11, no. 11, pp. 10571–10585, Nov. 2011.
 - [18] C. Yam, M. Nixon, and J. Carter, “Gait recognition by walking and running: a model-based approach,” *Asian Conf. Comput. Vis.*, no. January, pp. 1–6, 2002.
 - [19] A. Świtoński, A. Polański, and K. Wojciechowski, “Human identification based on gait paths,” in *International Conference on Advanced Concepts for Intelligent Vision Systems*, 2011, pp. 531–542.
 - [20] D. Cunado, M. S. Nixon, and J. N. Carter, “Automatic extraction and description of human gait models for recognition purposes,” *Comput. Vis. Image Underst.*, vol. 90, no. 1, pp. 1–41, 2003.
 - [21] J. A. Zeni, J. G. Richards, and J. S. Higginson, “Two simple methods for determining gait events during treadmill and overground walking using kinematic data,” *Gait Posture*, vol. 27, no. 4, pp. 710–714, May 2008.
 - [22] I. Bouchrika and M. S. Nixon, “Model-Based Feature Extraction for Gait Analysis and Recognition,” *Comput. Vision/Computer Graph. Collab. Tech.*, pp. 150–160, 2007.
 - [23] S. U. Jung and M. S. Nixon, “Heel strike detection based on human walking movement for surveillance analysis,” *Pattern Recognit. Lett.*, vol. 34, no. 8, pp. 895–902, Jun. 2013.
 - [24] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, “A database and evaluation methodology for optical flow,” in *International Journal of Computer Vision*, 2011, vol. 92, no. 1, pp. 1–31.
 - [25] C. Sun, “Fast Optical Flow Using Cross Correlation and Shortest-Path Techniques,” in *Proceedings of Digital Image Computing: Techniques and Applications*, 1999, pp. 143–148.
 - [26] H. Zimmer *et al.*, “Complementary optic flow,” in *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2009, pp. 207–220.
 - [27] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, “High accuracy optical flow estimation based on a theory for warping,” in *European conference on computer vision*, 2004, pp. 25–

36.

- [28] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1385–1392.
- [29] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, "SIFT flow: Dense correspondence across different scenes," in *European Conference on Computer Vision*, 2008, pp. 28–42.
- [30] W. Trobin, T. Pock, D. Cremers, and H. Bischof, "An unbiased second-order prior for high-accuracy motion estimation," in *Joint Pattern Recognition Symposium*, 2008, pp. 396–405.
- [31] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, "An improved algorithm for TV-L1 optical flow," *Stat. Geom. Approaches to Vis. Motion Anal.*, pp. 23–45, 2009.
- [32] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume." pp. 8934–8943, 2018.
- [33] A. Dosovitskiy *et al.*, "FlowNet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2758–2766.
- [34] Y. Sun, J. S. Hare, and M. S. Nixon, "Detecting acceleration for gait and crime scene analysis," in *Imaging for Crime Detection and Prevention (ICDP)*, 2016, no. 6, pp. 1–6.
- [35] C. Chen, Y. Shao, and X. Bi, "Detection of Anomalous Crowd Behavior Based on the Acceleration Feature," *IEEE Sens. J.*, vol. 15, no. 12, pp. 7252–7261, 2015.
- [36] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Imaging*, vol. 130, pp. 674–679, 1981.
- [37] Z. Dong, J. Qin, and Y. Wang, "Multi-stream deep networks for person to person violence detection in videos," in *Chinese Conference on Pattern Recognition. Springer*, 2016, pp. 517–531.
- [38] B. K. B. Horn and B. G. Schunck, "Determining Optical Flow," *Artif. Intell.*, vol. 17, no. 1–3, pp. 185–203, 1981.
- [39] D. Eager, A. M. Pendrill, and N. Reistad, "Beyond velocity and acceleration: Jerk, snap and higher derivatives," *Eur. J. Phys.*, vol. 37, no. 6, Nov. 2016.
- [40] R. Kalsoom and Z. Halim, "Clustering the driving features based on data streams," in *Inmic*, 2013, pp. 89–94.
- [41] Y. L. Murphey, R. Milton, and L. Kiliaris, "Driver's style classification using jerk analysis," in *2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems*, 2009, pp. 23–28.
- [42] O. Bagdadi and A. Várhelyi, "Development of a method for detecting jerks in safety critical events," *Accid. Anal. Prev.*, vol. 50, pp. 83–91, 2013.
- [43] B. Bringmann and P. Maglie, "A method for direct evaluation of the dynamic 3D path accuracy of NC machine tools," *CIRP Ann. Technol.*, vol. 58, no. 1, pp. 343–346, Jan. 2009.
- [44] M. P. Caligiuri, H. L. Teulings, J. V. Filoteo, D. Song, and J. B. Lohr, "Quantitative measurement of handwriting in the assessment of drug-induced parkinsonism," *Hum. Mov. Sci.*, vol. 25, no. 4–5, pp. 510–522, Oct. 2006.
- [45] L. Kong, X. Li, G. Cui, W. Yi, and Y. Yang, "Coherent Integration Algorithm for a Maneuvering Target with High-Order Range Migration," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4474–4486, Sep. 2015.
- [46] J. Zhang, T. Su, J. Zheng, and X. He, "Novel Fast Coherent Detection Algorithm for Radar Maneuvering Target with Jerk Motion," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 5, pp. 1792–1803, May 2017.

- [47] A. Datta, M. Shah, and N. Da Vitoria Lobo, "Person-on-person violence detection in video data," *Proc. Int. Conf. Pattern Recognit.*, vol. 16, no. 1, pp. 433–438, 2002.
- [48] M. Zaki, T. Sayed, and K. Shaaban, "Use of Drivers' Jerk Profiles in Computer Vision-Based Traffic Safety Evaluations," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2434, pp. 103–112, 2014.
- [49] N. Saunier and T. Sayed, "A feature-based tracking algorithm for vehicles in intersections," in *The 3rd Canadian Conference on Computer and Robot Vision (CRV'06)*, 2006.
- [50] L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1744–1757, 2012.
- [51] T. Komura, A. Nagano, H. Leung, and Y. Shinagawa, "Simulating pathological gait using the enhanced linear inverted pendulum model," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1502–1513, 2005.
- [52] S. Kajita, O. Matsumoto, and M. Saigo, "Real-time 3D walking pattern generation for a biped robot with telescopic legs," *Proc. 2001 ICRA. IEEE Int. Conf. Robot. Autom.*, vol. 3, pp. 2299–2306, 2001.
- [53] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter, "On a Large Sequence-Based Human Gait Database," *Appl. Sci. Soft Comput.*, pp. 339–346, 2013.
- [54] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 5, pp. 1511–1521, Oct. 2012.
- [55] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition," *IPSP Trans. Comput. Vis. Appl.*, vol. 10, no. 1, p. 4, 2018.
- [56] E. B. Nievas, O. D. Suarez, G. B. Garcia, and R. Sukthankar, "Violence Detection in Video Using Computer Vision Techniques," *Int. Conf. Comput. Anal. images patterns*, pp. 332–339, 2011.
- [57] A. Edison and C. Jij, "Optical Acceleration for Motion Description in Videos," *In Proceeding of CVPR Workshop*, pp. 39–47, 2017.
- [58] A. Edison and C. Jij, "Automated Video Analysis for Action Recognition using Descriptors Derived from Optical Acceleration", *Signal, Image and Video Processing*, vol. 13, pp. 915–922, 2019.
- [59] L. Wang, Y. Qiao and X. Tang, "Action Recognition with Trajectory-Pooled Deep-Convolutional Descriptors", *In Proceeding of CVPR*, pp. 4305–4314, 2015.
- [60] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features", in *Proc. IEEE Int. Workshop Vis. Surveill. Perform. Eval. Tracking Surveill.*, pp. 65–72, 2005.