Thesis: Yan Sun (2018) "On Classification of Acceleration and its Components in Computer Vision for Heel Strike Detection", University of Southampton, School of Electronic and Computer Science, PhD Thesis, pagination.

UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

SCHOOL OF ELECTRONIC AND COMPUTER SCIENCE

Vison, Learning and Control

On Classification of Acceleration and its Components in Computer Vision for Heel Strike Detection

by

Yan Sun

A thesis submitted in partial fulfilment for the
degree of Doctor of Philosophy

November 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

SCHOOL OF ELECTRONIC AND COMPUTER SCIENCE

Thesis for the degree of Doctor of Philosophy

On Classification of Acceleration and its Components in Computer Vision for Heel Strike Detection

Yan Sun

In some forms of gait analysis, it is important to be able to localise the heel within the frame in which the strike occurs. According to the motion characteristics of heel strikes, radial acceleration is ideal for estimating the spatio-temporal position of heel strikes in standard image sequences. Previous research of motion analysis has generally not yet considered the basic nature of higher orders of motion such as acceleration. Hence, in this thesis, acceleration first is computed in a principled manner by extending Horn and Schunck's algorithm for global optical flow estimation. We then demonstrate an approximation of the acceleration field using an alternative established optical flow technique, since most motion in real world violate the global smoothness assumption. Further, we decompose acceleration into radial and tangential based on geometry.

Compared with previous heel strike detection techniques, acceleration not only improves the precision significantly but also enables detection in real-time. Our new method also shows a good robustness in performance analysis with respect to noised image and occlusion. Acceleration is propagated as a general motion descriptor, it shows the capability for differentiating different types of motion both on synthesised data and real image sequences.

Beyond acceleration, the higher-orders of motion flow and their continuant parts are preliminarily investigated for further revealing the chaotic motion fields. Naturally it is possible to extend this notion further: to detect higher orders of image motion. In this respect we show how jerk and snap can be obtained from image sequences. The derived results on test images and heel strike detection

illustrate the ability of higher-order motion, which provide the basis for the following research and applications in the future.

Table of Contents

Table of Tables

Table of Figures

Academic Thesis: Declaration of Authorship

I, ... Yan Sun

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Motion Classification in Computer Vision ....

....

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

1) Y. Sun, J. S. Hare and M. S. Nixon, ‘Detecting Heel Strikes for Gait Analysis through Acceleration Analysis’, *IET Computer Vision*. vol. 12, no. 5, pp. 686 – 692, 2018.
2) Y. Sun, J. S. Hare and M. S. Nixon, ‘Analysing Acceleration for Motion Analysis’, in *International Conference on Signal Image Technology & Internet Based Systems*, Jaipur, India, 2017, pp. 289 – 295.
3) Y. Sun, J. S. Hare and M. S. Nixon, ‘Detecting Acceleration for Gait and Crimes Scene Analysis’, in *International Conference on Imaging for Crime Detection and Prevention,* Madrid, Spain, 2016, pp. 1-6.

Signed: ....

Date: ....

Acknowledgements

First, I would like to express my sincere gratitude to my supervisors Dr. Jonathon S. Hare and Professor Mark S. Nixon not only for their innovative advices and academic guidance, but also their encouragement for building up my confidence on research. I feel very lucky to have them as my supervisors in the past three and half years, I surely enjoyed and learnt a lot.

This work would not have been possible without the help from Osaka University for providing the OU-ISIR database and Imed Bouchrika from the University of Souk Ahras who provided the implementation of his algorithm for heel strike detection by corners. I am also grateful to The Institute of Automation, Chinese Academy of Sciences for making their databases CASIA publicly available.

I also want to thank my kindness colleagues in Vision, Learning and Control Group for their company, friendship and encouragements. We have together shared so much joy and laughter.

In the end, I want to delicate this thesis to my parents and my sister, it would not have been possible without their advices, support and love. Last, I hope my sister will enjoy her PhD as much as I did!

Definitions and Abbreviations

AE	Angular Error
AEPE	Average End-Point Error
CASIA	The CASIA Gait Database
EPE	End-Point Error
GT	Ground Truth
k-NN	k-Nearest Neighbours Algorithm
OU-ISIR	The OU-ISIR Gait Database
PR	Precise-Recall
ROI	Region of Interest
SD	Standard Deviations
SIFT	Scale-Invariant Feature Transform
SOTON	The Large Gait Database

Chapter 1 Introduction

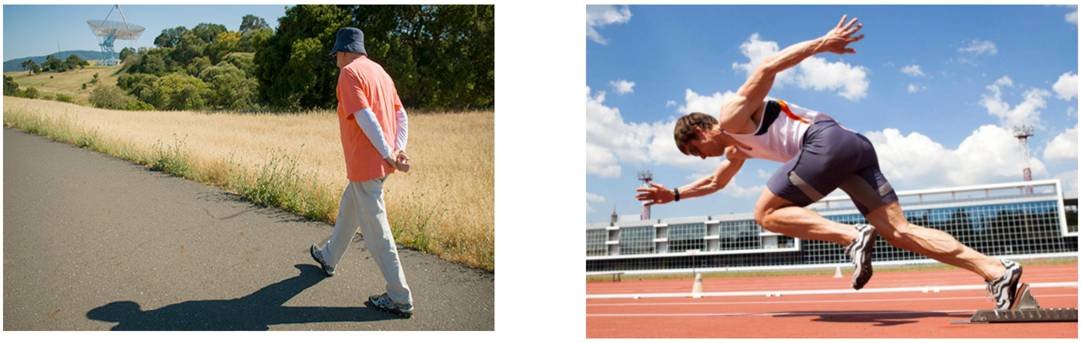
1.1 Motivation

Gait, the manner in which a persons' body moves when walking, is periodic. Most analysis methods rely on accurate gait phase detection, for example heel strike, foot stance and swing [1], [2]. It can be important to be able to capture when heel strikes occur in gait analysis; this is especially true in model-based recognition and forensic identification [3]. In addition, in terms of the video analysis of gait, localising where the heel strikes on the floor is essential since the stride and step length can be derived from the position of the heel [4]. A heel strike refers to the moment that the heel first strikes on the floor during walking. When the heel approaches the strike, the foot has significant radial acceleration which is centred at the heel. According to this inherent motion feature, radial acceleration is perfect for detecting the spatio-temporal positions of the heel strikes. There have been a number of physics-based gait analysis approaches using accelerometers and gyroscopes to detect the acceleration and angular velocity, to determine walking phases [1], [3].

Previous image-based heel strike detection methods usually accumulate the gait sequences and find the points where have the most desired features since the striking foot is stationary for almost half a period during one gait cycle [4], [5]. In addition, these techniques require the whole image sequence to determine the time and location, leaving detection in real-time an open problem. To extend heel strike detection to standard image sequences and enable near real-time analysis, a new generic motion descriptor based on optical-flow based estimation of acceleration fields, *acceleration flow*, is introduced in this thesis.

For heel strike detection, we propose an algorithm to classify motion within image sequences and subsequently decompose it into radial and tangential components. The key frame of heel strike can be determined by the quantity of acceleration flow within the Region of Interest (ROI), and the position of a heel strike can be found from the accumulated rotation centres caused by radial acceleration. Our approach has been tested on a number of datasets which were recorded indoors and outdoors with multiple views and walking directions for evaluating the detection rate under various environments.

The results of heel strike detection system suggest that acceleration is a distinctive motion feature and worthy of further investigation as a baseline approach. Nowadays, computer vision approaches can differentiate objects in motion from those which are static, but little more [6], [7]. An image is a snapshot in which all motions are frozen in time. This implies that video involves many motions which coalesce to form the image sequence. In reality, there are many different types of motion: in the simplest sense, there are objects that move with constant velocity and some that move with acceleration; however, in reality, many objects have more complicated motions. This thesis is the first systematic research aimed to disambiguate different levels of motion.



(a) A walking man

(b) A dasher

Figure 1.1 Different types of motion[1].

Figure 1.1 illustrates the diversity of motion, the man in the left image is walking with constant velocity in general and the athlete on the right is speeding up, or, accelerating. Moreover, each part of both subjects is experiencing different types of motion, especially the legs. When a person is walking the body moves at approximately a constant velocity, and one of the legs is stationary to support the body while the other one is swinging forward like a pendulum, as shown in Figure 1.2. These motions can be identified by acceleration as once the status of an object has changed, there must be acceleration. Therefore, we hypothesise that we can find the legs of a person's body and discriminate the supporting leg and the swinging one by extracting their acceleration features.

Figure 1.2 A walking cycle [8].

Subsequently, we explore the higher orders of motion and their constituent parts in the flow fields. The experimental results of a synthetic and real-world test images provide different characters among acceleration, jerk and snap. Then jerk and snap are applied in heel strikes detection again,

---

[1] Images are taken from: http://news.stanford.edu/2014/04/24/walking-vs-sitting-042414/ and http://www.wisegeek.com/what-are-the-different-types-of-track-spikes.htm#man-running-on-stret

the experimental results demonstrate the algorithm is ready for developing new applications in computer vision.

1.2 Contributions

Within this thesis, we have made the following contributions:

- We extend the original Horn-Schunck optical flow technique to focus on acceleration. Our new analysis retains the elegance of Horn and Schunck's formulation with an approach that isolates only acceleration.
- Our analysis shows the constraints within the algorithm to be too stringent for application in real-world video footage, so we explore the use of the other state of art optical flow algorithms as a basis for approximating acceleration with wider applicability in general video.
- Acceleration is decomposed into its constituent parts: radial and tangential acceleration.
- We use radial acceleration to localize the frame and position of heel strikes for gait analysis. This new method only needs three frames to determine the event compared with previous techniques which need the whole sequence.
- The experimental results show that our method can increase the precision of heel strike location significantly, especially when combined with simple classification methods like mean shift.
- The sensitivity of our approach to different imaging conditions is evaluated via a wide range of datasets, as well as different types of distortion: visual angle, lighting condition, Gaussian noise, occlusion and low resolution.
- Compared with other heel strike detection techniques, radial acceleration is less sensitive to Gaussian noise noticeably among the three types of noise, which would probably appear in real CCTV footages, whereas more sensitive to the occlusion in the detection region.
- The change of acceleration, jerk and snap flow, and their constituent parts are preliminary investigated on both synthetic and real images, the results show the potential for further application and study.

1.3 Publications

The following publications have resulted from the research presented in this thesis:

- Y. Sun, J. S. Hare and M. S. Nixon, 'Detecting Heel Strikes for Gait Analysis through Acceleration Analysis', *IET Computer Vision*. vol. 12, no. 5, pp. 686 – 692, 2018.

- Y. Sun, J. S. Hare and M. S. Nixon, ‘Analysing Acceleration for Motion Analysis’, in *International Conference on Signal Image Technology & Internet Based Systems*, Jaipur, India, 2017, pp. 289 – 295.
- Y. Sun, J. S. Hare and M. S. Nixon, ‘Detecting Acceleration for Gait and Crimes Scene Analysis’, in *International Conference on Imaging for Crime Detection and Prevention,* Madrid, Spain, 2016, pp. 1-6.
- Y. Sun, J. S. Hare and M. S. Nixon, On Parameterizing Higher-order motion for behaviour recognition. (To be submitted)

1.4 Thesis Outline

This thesis is arranged as follows: Chapter 2 gives a brief introduction of gait analysis and optical flow, four benchmark algorithms are compared to demonstrate the advantages and weakness of optical flow algorithms. Our acceleration algorithms and the experimental results on both synthetic and real-world images are presented in Chapter 3. Chapter 4 describes and evaluates the methodology of detecting heel strike via radial acceleration. In Chapter 5, optical flow is decomposed into higher orders and their constituent parts, the analysis and results are ripe for further investigation. Chapter 6 concludes our work and explores potential future directions.

Chapter 2 A Review of the State of the Art

2.1 Introduction

The heart of this thesis is centred around the idea of using acceleration as a feature for detecting heel strikes. In this thesis the acceleration algorithm is built based on optical flow. This chapter first introduces gait, and we stress the importance of heel strike in the analysis also briefly review the state of the art techniques. The second part reviews optical flow algorithms, then introduces the principles of some representative optical flow algorithms and compares their performances on synthetic and real images.

This chapter is arranged as follows: Section 2.2 introduces gait analysis and previous heel strike detection algorithms. Section 2.3 gives a general review of optical flow. Section 2.4 describes the most classical techniques. Section 2.5 introduces the synthetic images for illustrating the simple flow fields, as well as the general visualization methods and error measurements of flow. The optical flow techniques are evaluated and discussed in Section 2.6. In the end, Section 2.7 concludes this chapter.

2.2 Gait Analysis and Heel Strike Detection

2.2.1 Gait Analysis

Gait analysis is the systematic study of human walking. It has been mainly applied in two fields: medical consultation for conditions which affect walking [9] and human identification [10]. Clinical gait analysis usually uses physical data to analyse the walking pattern of patients for diagnosis and treatment. The data is collected by wearable or non-wearable sensors, like accelerometer and treadmill.

In human identification, as a behavioural biometric which obtained at a distance from the camera gait is hard to hide or disguise. It is the most reliable biometric in the criminal investigation since it is less sensitive to the low quality of images compared with other biometrics. It has been demonstrated that gait can be used in criminal investigations either as the body [11] or the gait measurements [12]. Figure 2.1 shows a CCTV footage of an Australian jewellery shop murder: the target covered his face during the crime, however he was recognised by his gait after it was found that he had come to the jewellery shop earlier that day.

Figure 2.1 Murder who was recognised by his gait.

The approaches to analysing gait can be classified as three types based on the sensor modalities that are used to make measurements: physical-sensor based, depth image based and standard image based. Physical-sensor based techniques measure the physical data extracted from gait, mostly kinetic parameters and underfoot force/pressures [25],[26].

The physical-sensors are classifiable as wearable and non-wearable. The most popular wearable sensors are accelerometers and gyroscopes [1], [13]. Milica et al. [14] use accelerometer to measure the angle of leg segments and ankle. In Rueterbories et al.'s work [15] they use gyroscopes to capture the angular displacement, or Coriolis force since it is the response to a rotating particle, to discriminate gait events. In the modalities of underfoot force/pressure sensors, researchers use the features of Ground Reaction Forces (GRFs) for analysing gait and it is commonly considered as the golden standard for gait phase partitioning [3]. In the recent research into GRF, Derlatka use Dynamic Time Warping (DTW) to measure the stride difference and then use k-Nearest Neighbour (k-NN) to classify people [16]. Later Derlatka and Bogdan partitioned the GRF stance into five sub phases to achieve a higher classification rate [17].

Depth, or RGBD image, based gait analysis techniques have expanded since the introduction and wide availability of PrimeSense and Kinect sensors. These measurements use the distance between the body parts and the sensor in depth images to analyse gait [18], [19]. Lu and et al. [20] have built a gait database named ADSC-AWD based on Kinect data. O'Connor measure the acceleration of the body using Kinect.

Standard image-based gait recognition has been extensively studied. Most approaches are targeted at recognition of individual humans, using gait as the biometric signature. The general framework usually consists of background subtraction, feature extraction and classification [21]. The approaches can be classified into two catalogues: model-based and model-free. Model-based approaches have an intimate relationship with the human body and its motion. Świtoński extracted the velocities and accelerations across the path of skeleton root element, feet, hands and head as the gait feature [22]. Yam et al. presented an analytical gait model which extract the angle of thigh and lower leg rotation without parameter selection [23]. Model-free approaches concentrate on the body shape or the motion of the entire gait process and thus could be used for the analysis of other moving shapes or mammals. Bobick and Davis [24] employ the motion-energy image and motion-history images of silhouette, Han and Bhanu [25] use the gait-energy image for recognition. Model-based methods are view-invariant and scale-invariant but the computation cost is relatively high and the approaches can be very sensitive to image quality. Model-free approaches are less sensitive to the image quality with lower computation cost though they are not intrinsically robust to variation in viewpoint and scale [21].

2.2.2 Heel Strikes Detection for Gait Analysis

Gait is periodic and most analysis methods relies on accurate gait periods detection. The components of one gait cycle are shown in Figure 2.2: a gait cycle is defined as the interval between two consecutive heel strikes of the same foot. 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.

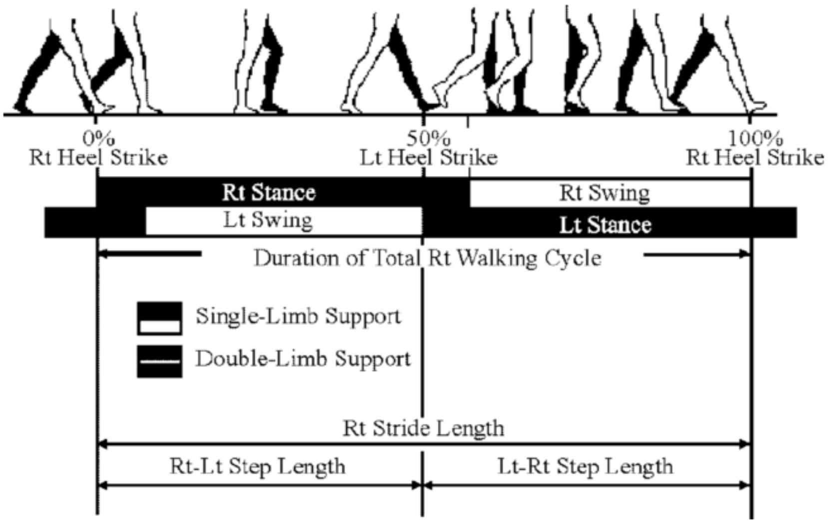


Figure 2.2 The temporal components contained in a gait cycle and step and stride length during the cycle [8].

Hence the accurate and efficient detection of heel strike is essential because it partitions walking sequences into cycles composed of stance and swing phases [26]. In addition, the stride and step length can be derived from the stationary position of the heel at the moment of strike. Heel strikes also show outstanding ability to disambiguate walking people from other moving objects.

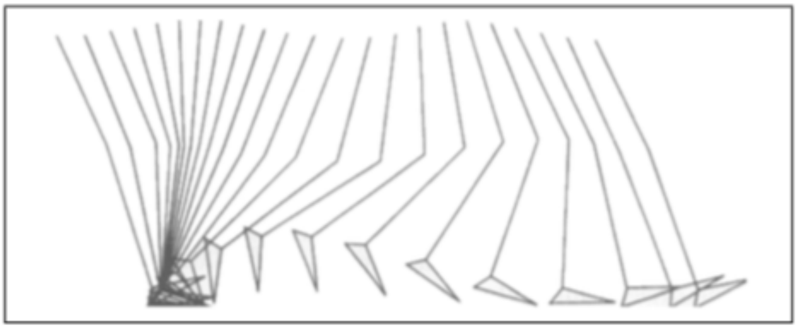
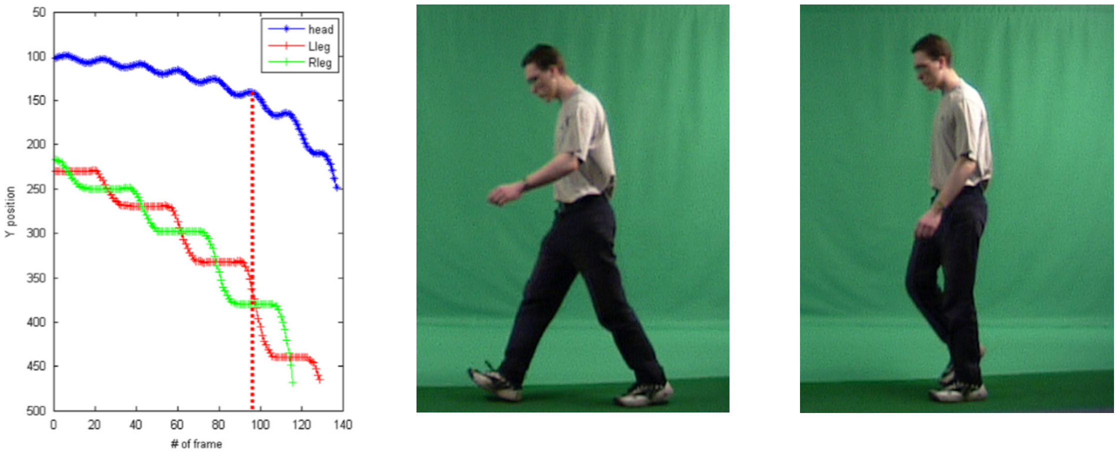


Figure 2.3 Foot model of walking [27].

The striking foot is stationary for almost half period during one gait cycle as shown in Figure 2.3. Therefore, previous standard image-based heel strike detection methods usually accumulate the gait sequences based on empirical analysis and find the areas where has the most desired features. Bouchrika and Nixon [28] extract low-level features, corners, to estimate the strike positions. They use the Harris corner detector to detect all the corners of each frame and obtain the corner proximity image by accumulate proximity algorithm. The positions of heel strike on the ground locate around the densest areas of accumulated corners.



(a) The vertical positions of the head, left leg and right leg [4]

(b) When the heel strikes

(c) When the feet are crossing

Figure 2.4 The relationship between the trajectories of head with gait events.

Jung and Nixon use the movement of the head to detect the key frame (the frame in which the heel strike takes place) [4]. When a person is walking, the vertical positions of the head in the sequence is similar to a sinusoid, shown in Figure 2.4 (a). When the heel strikes, the stride length is maximal, and so the head is at the lowest point; when the feet are crossing, the head is at the highest point. Similar to Bouchrika's method, Jung and Nixon accumulate the silhouettes of the whole sequence to find the positions where they remain the longest, as shown in Figure 2.5.

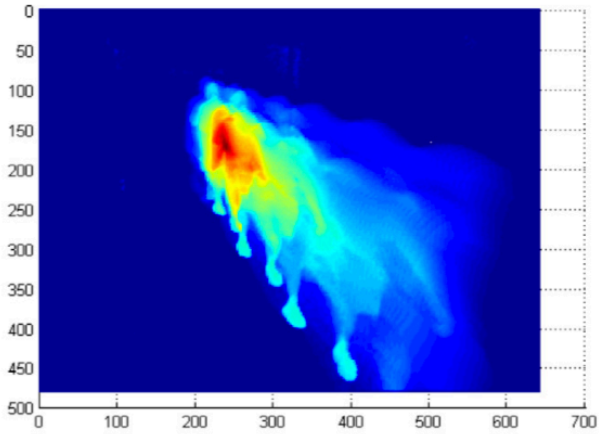


Figure 2.5 Silhouette accumulation map for a gait sequence [4].

2.3 A Review of Optical Flow algorithms

The concept of optical flow was first described by James J. Gibson in 1950. Optical flow denotes the apparent motion between the observer and the observed object caused by relative motion [6]. It has been widely used in many fields of image processing such as motion estimation and video compression. For an image, optical flow is the change of brightness patterns. Figure 2.6 (a) and (b) are two successive frames and (c) is the optical flow between them. Thus, many aspects of scene motion can be determined from optical flow: the people and the train are highlighted by optical flow whereas the static objects (e.g. the trees) are not.

(a) Frame n

(b) Frame $n+1$

(c) Optical flow field

Figure 2.6 The optical flow between two consecutive frames[2].

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 of optical flow did not change too much [6]. Most state-of-art approaches estimating optical flow by optimizing the weighted sum of two terms: a data term and a prior term [7]. Mathematically:

$$\mathbf{I} = \mathbf{I}_{data} + \alpha \mathbf{I}_{prior} \tag{2.1}$$

2.3.1 Data Term

2.3.1.1 Brightness Constancy

The fundamental assumption in optical flow is Brightness Constancy [29], assuming that the intensity at position (x, y) in frame t is $\mathbf{I}(x, y, t)$, the constraint is that the intensity of this point is constant between successive moment:

$$\mathbf{I}(x, y, t) = \mathbf{I}(x + \Delta x, y + \Delta y, t + \Delta t) \tag{2.2}$$

the change of the intensity can be expanded by Taylor series approximation:

$$\mathbf{I}(x + \Delta x, y + \Delta y, t + \Delta t) = \mathbf{I}(x, y, t) + \frac{\partial \mathbf{I}}{\partial x}\Delta x + \frac{\partial \mathbf{I}}{\partial y}\Delta y + \frac{\partial \mathbf{I}}{\partial t}\Delta t + O(\Delta i^2) \tag{2.3}$$

If $\Delta t \to 0$,

---

[2] Video is taken from: http://www.vision.ee.ethz.ch/datasets_extra

$$\frac{\partial \mathbf{I}}{\partial x}dx + \frac{\partial \mathbf{I}}{\partial y}dy + \frac{\partial \mathbf{I}}{\partial t}dt = 0 \tag{2.4}$$

After dividing by dt:

$$\frac{\partial \mathbf{I}}{\partial x}\frac{dx}{dt} + \frac{\partial \mathbf{I}}{\partial y}\frac{dy}{dt} + \frac{\partial \mathbf{I}}{\partial t} = 0 \tag{2.5}$$

If using u and v to denote $\frac{dx}{dt}$ and $\frac{dy}{dt}$ separately, gives the Optical Flow Constraint equation:

$$\left(\mathbf{I}_x, \mathbf{I}_y\right) \cdot (u, v) = -\mathbf{I}_t \tag{2.6}$$

There are two unknowns, u and v, with only one constraint. Therefore, in order to solve the problem, the prior term need to be introduced to the model (will be introduced in 2.3.2). The alternatives of data term are the correlation between frames and colour space. Sun [30] use fast cross-correlation to compute dense flow and Zimmer and et al. incorporate HSV in their data term [31].

2.3.1.2 Penalty Function

For estimating the flow, one key step is the penalty function to the violations. The most common choice is L2 norm [29], which simplifies the computation:

$$e_c = \iint \left(\mathbf{I}_x u + \mathbf{I}_y v + \mathbf{I}_t\right)^2 dxdy \tag{2.7}$$

Equation (2.7) corresponds to Gaussian assumption therefore it is not robust on the boundaries if there is occlusion. Black and Anandan [32] propose a framework based on robust estimation which is adapted to some later work [33], [34]; another popular penalty function is L1 norm [33], [35].

2.3.1.3 Other Features

Apart from the intensity of the frames, robust pairwise features also can be used for constructing the motion fields. Brox and et al. combine gradient constancy with brightness [35]; DeepFlow [36] and SIFT flow [37] both use Scale-Invariant Feature Transform (SIFT) for matching which performs the best among the illumination invariant features.

2.3.2 Prior term

To make the problem well-posed, an additional constraint needs to be introduced to the algorithm. The most widely used prior term is smoothness, which assumes that the neighbouring pixels (which belong to the same object) have a similar motion [29], [38]. If using L2 norm, the penalty function is:

$$e_s = \iint \left(\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial v}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2 \right) dxdy \tag{2.8}$$

Besides first order, Trobin and et al. use second-order prior to achieve high-accuracy optical flow [39], Wedel and et al. adapt rigid motion as their prior [33].

2.3.3 Learning Methods

It is hard to ignore learning algorithms in the upcoming optical flow approaches. Sun and et al. [40] learn a statistical model of both the error of brightness and smoothness constrains. FlowNet [41] uses convolutional neural networks (CNNs) to predict optical flow from a large quantity of training data.

Apart from the above review, numbers of optimization improve the performance of optical flow algorithms, however this is beyond the scope of this thesis. In the next section, we will give details and compare the performance of four benchmark algorithms among different types.

2.4 Selected Optical Flow Algorithms

2.4.1 Differential Method

Horn and Schunck developed the first differential approach to computing optical flow in 1981. It represented the beginning of variational techniques in Computer Vision [42]. Nowadays most upcoming algorithms still build their algorithms based on Horn-Schunck's theory. They estimated optical flow from the spatial-temporal derivatives of image intensity based on brightness constancy and motion smoothness, which have been introduced in 2.3. Combing Equation (2.7) and (2.8), the problem becomes one of minimizing the change of the optical flow along both the horizontal and vertical directions:

$$e = \alpha e_s + e_c = \iint \left(\alpha(\nabla^2 u + \nabla^2 v) + \left(\mathbf{I}_x u + \mathbf{I}_y v + \mathbf{I}_t\right)^2 \right) dxdy \tag{2.9}$$

where α is the factor of motion smoothness. The solution, the velocities (u, v), is obtained by minimizing the total error [29].

2.4.2 Region Based Method

Block Matching is one of the most fundamental methods in region based matching techniques. The algorithm assumes that the intensity of every single pixel remains constant between the successive frames if the motion is continuous (there is no occlusion) [6]. Optical flow can be easily computed by determining which block best matches the current block in a selected neighbourhood.

The implementation of a region based matching technique can be achieved by minimizing the sum-of-squared difference (SSD) between blocks in the image [43]:

$$SSD = \sum_{(x,y)\epsilon B} \left(\mathbf{I}(x + \Delta x, y + \Delta y, t + \Delta t) - \mathbf{I}(x, y, t)\right)^2 \tag{2.10}$$

where $\mathbf{I}(x, y, t)$ is the pixel intensity of position (x, y) at frame t. The matching block is that where the error is minimum within the search area, and the optical flow is therefore the change in position between the current block and the matched block [44].

2.4.3 Dense Optical Flow

Farneback developed the most popular dense optical flow algorithm based on polynomial expansion [45]. In the algorithm, the neighbour of each pixel can be approximated by polynomial expansion:

$$f(\mathbf{x}) \sim \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c \tag{2.11}$$

where $\mathbf{x} = (x \quad y)^T$, $\mathbf{A}$ is a 2x2 matrix and $\mathbf{b}$ is a 2x1 vector. Optical flow assumes that the image intensity is constant. Therefore, if the displacement between $f_1(x)$ and $f_2(\mathbf{x})$ is $\mathbf{d}$:

$$f_2(\mathbf{x}) = f_1(\mathbf{x} - \mathbf{d}) = (\mathbf{x} - \mathbf{d})^T \mathbf{A}_1 (\mathbf{x} - \mathbf{d}) + \mathbf{b}_1^T (\mathbf{x} - \mathbf{d}) + c_1 = \mathbf{x}^T \mathbf{A}_2 \mathbf{x} + \mathbf{b}_2^T \mathbf{x} + c_2 \tag{2.12}$$

$$\begin{cases} \mathbf{A}_2 = \mathbf{A}_1 \\ \mathbf{b}_2 = \mathbf{b}_1 - 2\mathbf{A}_1 \mathbf{d} \\ c_2 = \mathbf{d}^T \mathbf{A}_1 \mathbf{d} - \mathbf{b}_1^T \mathbf{d} + c_1 \end{cases} \tag{2.13}$$

Thus if $\mathbf{A}_1$ is non-singular, the displacement $\mathbf{d}$ is:

$$\mathbf{d} = -\frac{1}{2}\mathbf{A}_1^{-1}(\mathbf{b}_2 - \mathbf{b}_1) \tag{2.14}$$

2.4.4 DeepFlow

DeepFlow has emerging as a popular optical flow technique in recent years due to its excellent performance on large displacement estimation and non-rigid matching and was developed by Weinazepfel and et al. in 2013. DeepFlow made a step towards bridging the gap between descriptor matching algorithms with large displacement optical flow techniques [36]. An outline of DeepFlow is shown in Figure 2.7. It will be introduced it in two parts: the deep matching algorithm and the energy minimization framework.

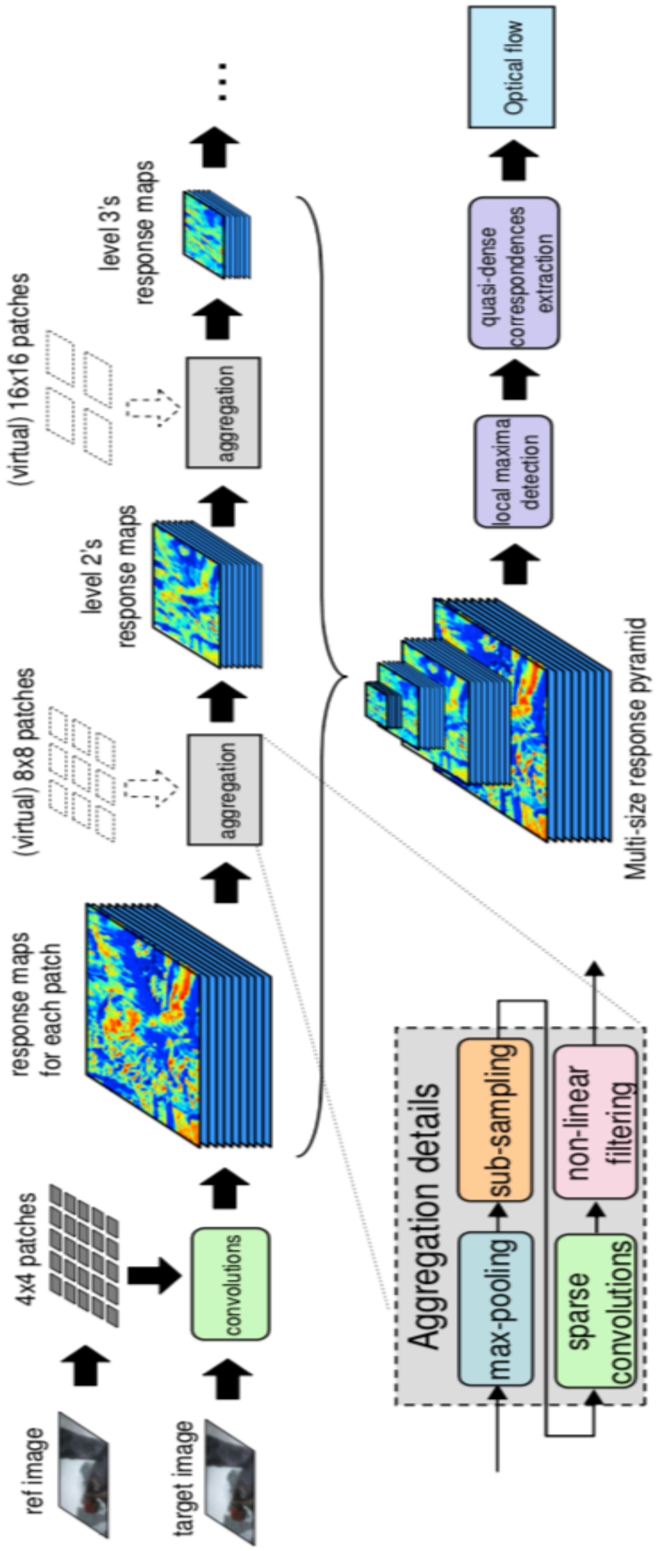


Figure 2.7 The outline of DeepFlow [36].

2.4.4.1 DeepMatching

The deep matching algorithm first splits the Scale-Invariant Feature Transform (SIFT) [46] descriptor from a 128-dimensional real vector into four quadrants: the gradient orientations of the interest point are changed from $H \in R^{128}$ into $H = [H^1 H^2 H^3 H^4]$ where $H^s \in R^{32}$. In order to maximize the similarity between the reference and target descriptor, DeepFlow optimizes the positions of H^s on target descriptor by assuming that each of the quadrants can moving independently with some extent rather than keep them fixed:

$$sim\big(H, Q(p)\big) = \sum_{s=1}^{4} max_{p_s} H_s^T Q(p_s) \tag{2.15}$$

where $Q(p) \in R^{32}$ is one quadrant of the reference descriptor. By assuming the quadrants are moving independently, a coarse-to-fine non-rigid matching can be obtained efficiently.

If $\{P_{i,j}\}_{i,j=0}^{L-1}$ and $\{P'_{i,j}\}_{i,j=0}^{L-1}$ denotes the reference and target descriptor respectively, the optimal warping ω^* is the one of that maximizes the similarity between the pixels:

$$S(\omega^*) = \max_{\omega \epsilon W} S(\omega) = \max_{\omega \epsilon W} \sum_{i,j} sim\Big(P(i,j), P'\big(\omega(i,j)\big)\Big) \tag{2.16}$$

where $\omega(i)$ returns the position of pixel i in P'. If define recursively then we can obtain the optimal warpings that are largely robust to deformation [36].

2.4.4.2 Energy Minimization Framework

DeepFlow is an energy minimization function which is similar to Horn-Schunck. It is based on the same two assumptions: intensity constancy and smooth motion. In addition, an extra term deep matching, is blended into the framework:

$$E(\omega) = \int E_D + \alpha E_s + \beta E_M \, dx \tag{2.17}$$

where E_D is the weighted data sum, E_S is the smoothness term and a matching term E_M. A robust penalizer is applied to each term:

$$\Psi(s) = \sqrt{s^2 + \epsilon^2} \tag{2.18}$$

with $\epsilon = 0.001$ which was determined empirically [47]. The data term consists of two penalizers of brightness:

$$E_D = \delta\Psi\left(\sum_{i=1}^{c} \omega^T \bar{J}_0^i\, \omega\right) + \gamma\psi\left(\sum_{i=1}^{c} \omega^T \bar{J}_{xy}^i\, \omega\right) \tag{2.19}$$

where the first term is the penalizer over image channels, the second one is the penalization for the x and y axes. ω is the flow we seek to estimate: $(u, v)^T$, c is the number of the image channels. $\bar{J_0}$ is a tensor which is normalized by spatial derivatives:

$$\bar{J}_0^i = \theta_0\left(\nabla_3 I^i\right)\left(\nabla_3^T I^i\right) \tag{2.20}$$

∇_3 is the spatial-temporal gradient $(\partial x, \partial y, \partial z)$. θ_0 is the spatial normalization factor $\left(\left\|\nabla_2 I^i\right\|^2 + \xi^2\right)^{-1}$ to reduce the impact of small gradient locations and $\xi = 0.1$ to prevent the factor to be zero. The gradient constancy penaliser is normalized along the x and y axes respectively:

$$\bar{J}_{xy}^i = \left(\nabla_3 I_x^i\right)\left(\nabla_3^T I_x^i\right)\left(\left\|\nabla_2 I_x^i\right\|^2 + \xi^2\right)^{-1} + \left(\nabla_3 I_y^i\right)\left(\nabla_3^T I_y^i\right)\left(\left\|\nabla_2 I_y^i\right\|^2 + \xi^2\right)^{-1} \tag{2.21}$$

where I_x and I_y are the gradient derivatives with respect to the horizontal and vertical axis. The smoothness term in DeepFlow is a penalization for gradient flow:

$$E_s = \Psi(\|\nabla u\|^2 + \|\nabla v\|^2) \tag{2.22}$$

The purpose of the matching term is to find the most similar flow to the known vector ω' as previously introduced in Section 2.4.4.1. The difference is estimated by:

$$E_M = b\phi\Psi\|\omega - \omega'\|^2 \tag{2.23}$$

Due to the matching being semi-dense, a binary term b(x) is added into the matching term. b(x) equals 1 if and only if there is a match at position x. $\phi(x)$ is a weight that is low in the flat area. The optical flow we seek to estimate $\omega: (u, v)^T$ can be obtained by minimising the energy function [36].

In this section, we introduced the principles of the classical optical flow algorithms from different categories as well as the state of art. Before the evaluation, it is worth describing the performance quantification. The performance of EpicFlow and FlowNet will not be included in the evaluation as these are not yet standard approaches nor do they perform better than DeepFlow in the public optical flow algorithm evaluation[3].

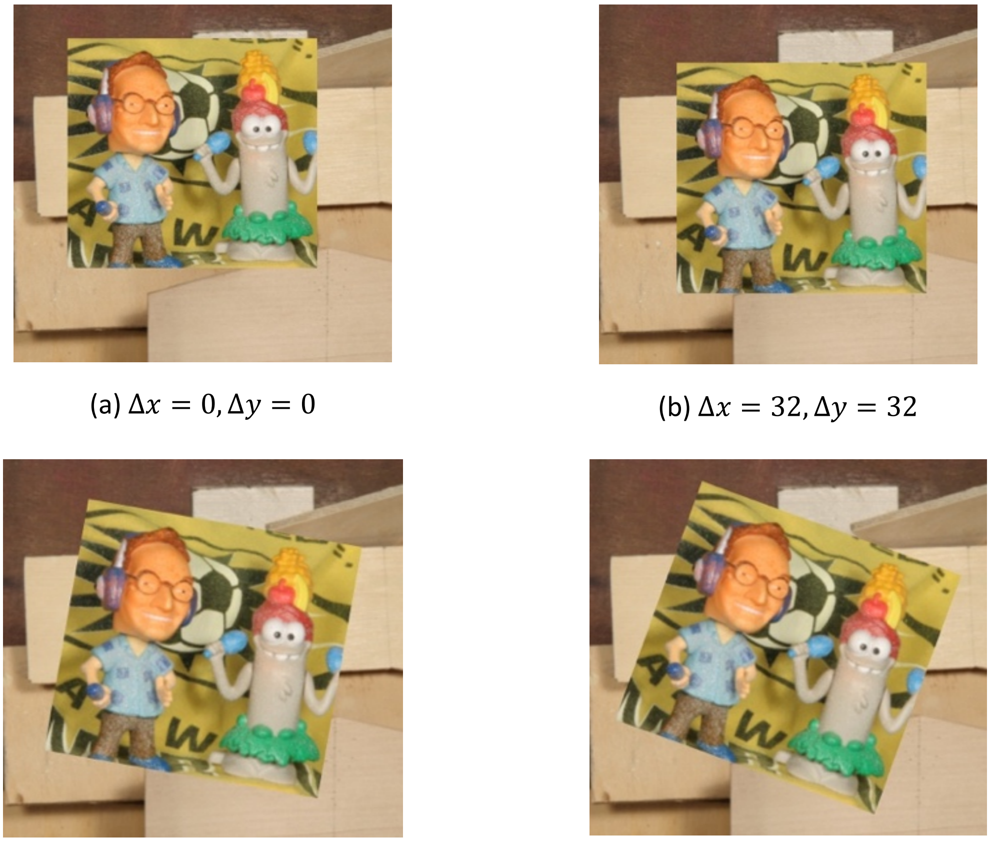
---

[3] http://vision.middlebury.edu/flow/eval/results/results-e1.php

2.5 Preparation for Performance Quantify of Optical Flow Algorithms

2.5.1 Synthetic Images with Explicit Motion

The advantage of synthetic images is that they lack specularity, or other types of noise. Also, the motion field and the scene properties can be manipulated as required. For evaluating whether the algorithms of acceleration and its components (will be introduced in Chapter 3) can estimate the desired flow, some test images with only simple motion (like linear shift or rotation) are necessary. Two sets of synthetic image sequences were constructed and help us to identify where the flow approaches fail in the first place. The artificial image sequences involve linear translation and rotation are synthesized by using images from the Middlebury database [7]. A subpart of a frame from Mequon (the block of two faces in the middle of synthetic images shown in Figure 2.8) in Middlebury is embedded in a frame from the Wooden images. The sub-Mequon shifts along a linear trajectory to the lower right corner at speed 1 and 3 pixels/frame, both on horizontal and vertical axes. The rotation sequence is obtained by rotating the middle square around its centre to form circular motion. Figure 2.8 gives the examples of linear shifting when the displacement is 32 pixels from the start position and rotation images when the rotation degrees are 10° and 30°.



(a) $\Delta x = 0, \Delta y = 0$

(b) $\Delta x = 32, \Delta y = 32$

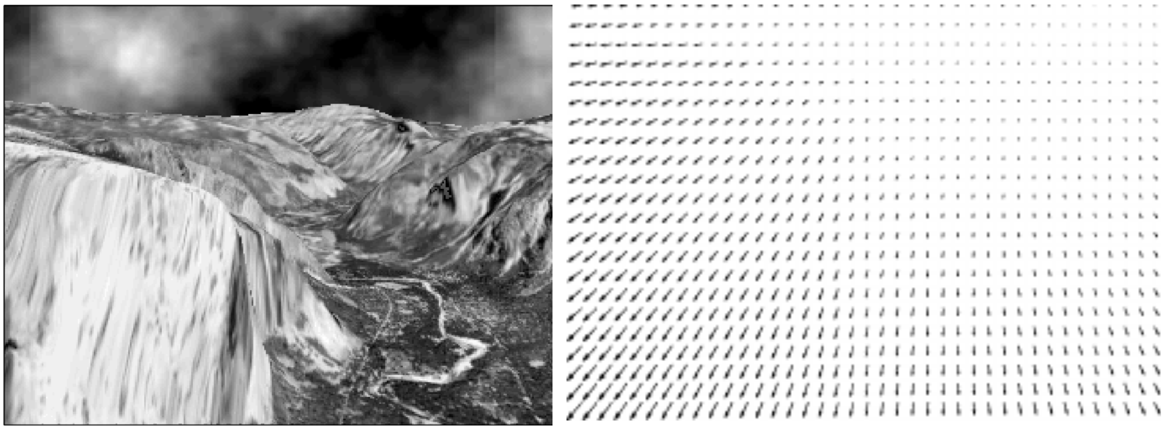
(c) $\Delta \alpha = 10°$

(d) $\Delta \alpha = 30°$

Figure 2.8 Examples of synthetic test images.

2.5.2 Flow Visualization

In the early optical flow papers, the flow field is visualized by an arrow which starts from the initial position points to the moving direction, the length of the arrow indicates the magnitude of the displacement. Figure 2.9 gives an example of using arrow to present the motion field.



(a) A frame of Yosemite sequence

(b) Motion field visualized by arrow

Figure 2.9 Presenting flow using arrow [45].

With the development of optical flow, the new approaches can handle more complicated and anisotropic situations, so using arrows might cause confusion in the analysis of a complex motion field. Baker and et al. [7] created a colour coding scheme for visualizing such complex motion when they established their optical flow dataset. The colour scheme is shown in Figure 2.10: the hue indicates the direction and the saturation represents the intensity of the flow.

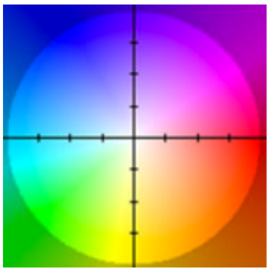


Figure 2.10 Flow field colour coding.

2.5.3 Flow Error Measurements

In general, there are two commonly used measurements in optical flow: Angular Error (AE) and End-Point Error (EPE) [1], [12]. If the ground-truth vector and estimated vector are $(u_{GT}, v_{GT})^T$, $(u, v)^T$, AE measures the angle between the ground-truth flow and predicted flow vector in a 3D space (pixel, pixel, frame):

$$AE = \cos^{-1}\left(\frac{1.0 + u \times u_{GT} + v \times v_{GT}}{\sqrt{1 + u^2 + v^2}\sqrt{1 + {u_{GT}}^2 + {v_{GT}}^2}}\right) \tag{2.24}$$

where $(u_{GT}, v_{GT}, 1)^T$, $(u, v, 1)^T$ are the extended 3D vectors. The second measurement, EPE, is the Euclidean distance between the two vectors on a 2D image plane:

$$EPE = \sqrt{(u - u_{GT})^2 + (v - v_{GT})^2} \tag{2.25}$$

Both of the measurements have their own advantages and bias: AE is more sensitive to the error of small motion whereas it undervalues the errors with large motion. EPE strongly penalizes the large motion errors but is insensitive to errors with small motion [9]. Now we have given a general sense of how the flow can be presented and evaluated, the next section is going to test the methods introduced in Section 2.4 and illustrate the results in the most appropriate way.

2.6 Experimental Results of Existing Algorithms on Synthetic Images

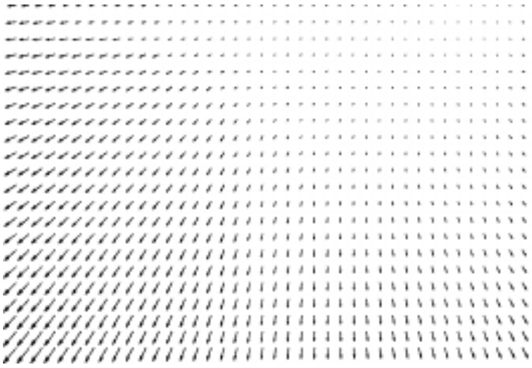
The optical flow methods are first tested on the synthetic linear translation sequence, the results produced from simple motion can highlight the drawbacks of techniques. The collection of flow fields is shown in Figure 2.11. Although the flow fields are self-evident, some discussion is still required.

Beginning with small motion it can be observed that all the techniques produce reasonably accurate flow fields. When the displacement increases to 3 pixels, Horn-Schunck and Block Matching have poor estimations since they are global algorithms without multi-layer refinement. Block Matching has more error on the smooth area since local method will be ill-posed in homogeneous region [48]. The accuracy largely relies on an adequate block size according to the motion and texture in the image. Farneback gives some directional error along the block edge. As the most advanced approach, DeepFlow performs the best on both scenarios, including the discontinuous area.

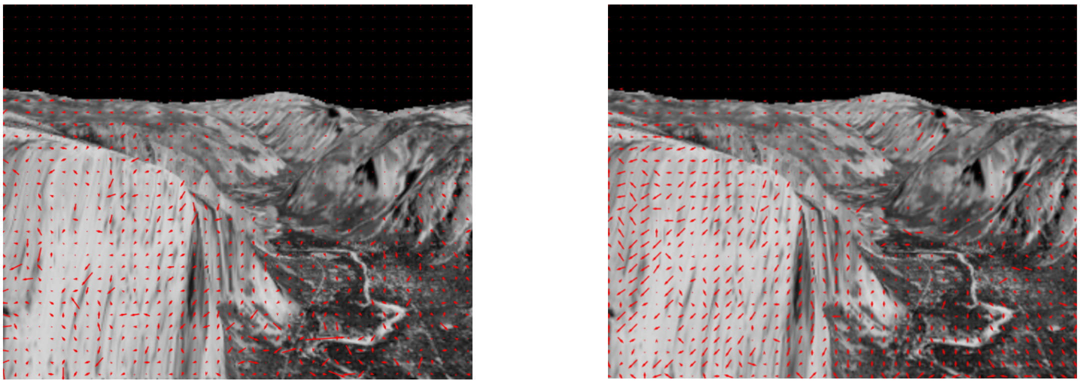

Small rigid motion (1 pixel)
Large rigid motion (3 pixels)
Horn-Schunck
Block Matching
Farneback

Figure 2.11 Estimated flow fields on small and large rigid motion.

In addition to the synthetic images, the optical flow methods are also evaluated on the famous test image sequence, Yosemite, shown in Figure 2.12. It is a conventional challenge for most optical flow algorithms due to the divergent displacement on different areas and the edges between the mountains are occluded as the scene moves. The non-uniform motion is caused by the asymmetrical projection of 3D motion in real-world onto the 2D image surface: the upper right corner translates to the right with a speed of 2 pixels/frame and the speed in the lower left area is about 4 to 5 pixels/frame [43]. Although the scene is complex, the motion field is simple. The camera is moving straight forward smoothly without any rotation or distortion. The GT is extracted from [49] since it is not publicly available to prevent overfitting by new optical flow algorithms.

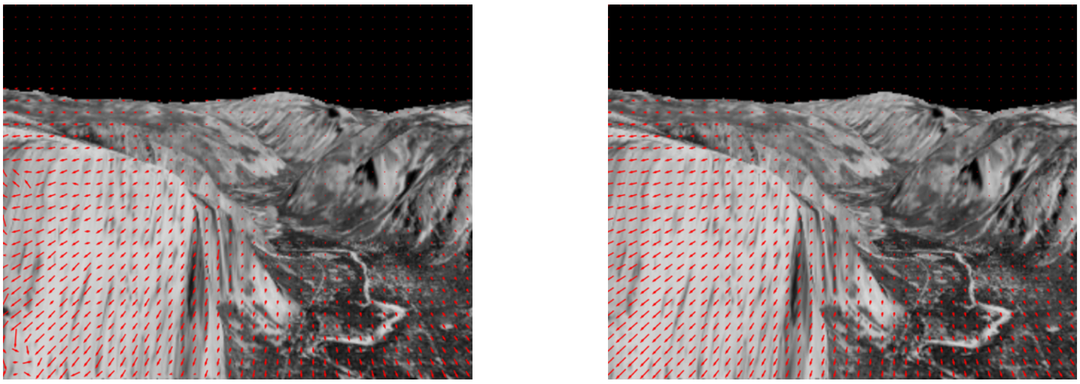


(a) GT of Yosemite



(b) Horn-Schunck

(c) Block Matching



(d) Farneback

(e) DeepFlow

Figure 2.12 The GT of Yosemite and the estimated flow fields.

The results are identical within synthetic sequences, the estimation result by DeepFlow is linear and evenly distributed along the moving direction of the camera as expected; Farneback is noisy at the edge of the image whereas the results of Horn-Schunck and Block Matching are poor in this test image.

For a more objective evaluation, we applied these algorithms on seven image sequences (GT is available) from Middlebury [7] to give a statistical measurements. Table 2.1 summaries the Average End-Point Error (AEPE), which is the sum of the distance between estimated flow and GT averaged over all the points with distance exceeds the threshold. Table 2.2 reports the Standard Deviations (SD) of the values shown in Table 2.1.

Method	Dimetrodon	Hydrangea	RubberWhale	Urban2	Urban3	Grove2	Grove3	Avg.
Horn-Schunck	1.66	3.35	0.61	8.06	7.18	3.16	3.89	3.99
Block Matching	1.78	2.79	0.82	8.09	7.05	1.37	2.98	3.55
Farneback	0.26	0.65	0.21	7.53	6.75	0.47	2.37	2.61
DeepFlow	0.11	0.17	0.13	0.29	0.44	0.18	0.66	0.28

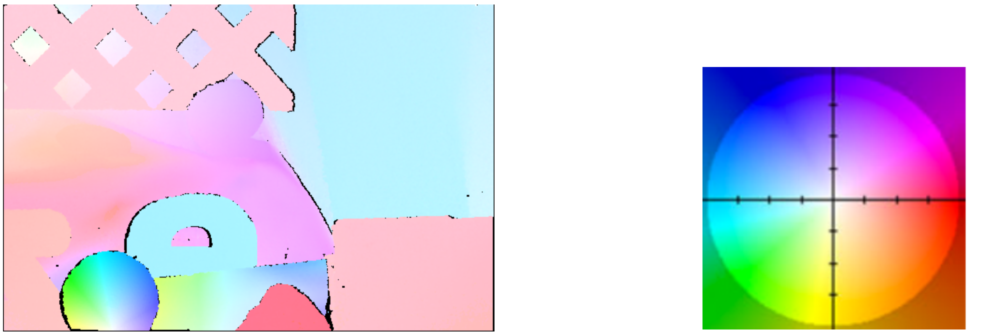
Table 2.1 AEPE on Middlebury datasets.

Method	Dimetrodon	Hydrangea	RubberWhale	Urban2	Urban3	Grove2	Grove3	Avg.
Horn-Schunck	0.94	1.6	0.65	8.17	4.83	1.32	2.96	2.92
Block Matching	1.24	2.13	0.88	8.23	4.85	1.57	2.93	3.12
Farneback	0.41	1.48	0.48	8.85	5.53	0.96	3.12	2.98
DeepFlow	0.1	0.36	0.26	0.95	1.44	0.43	1.45	0.71

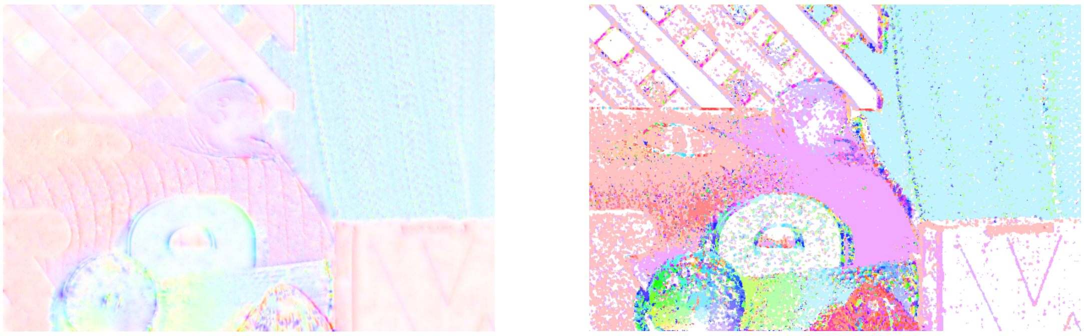
Table 2.2 The SD of AEPE.

Furthermore, Figure 2.13 and Figure 2.14 exhibit the distribution of the measurements of RubberWhale and Dimetrodon and support Table 2.1 and Table 2.2. Different from the manipulated image sequences, the motion fields in Middlebury are more complex than the synthetic sequence, we opt to use colour coding to illustrate the flow field.

(a) Frame 10 (b) Frame 11



(c) GT (d) Colour coding

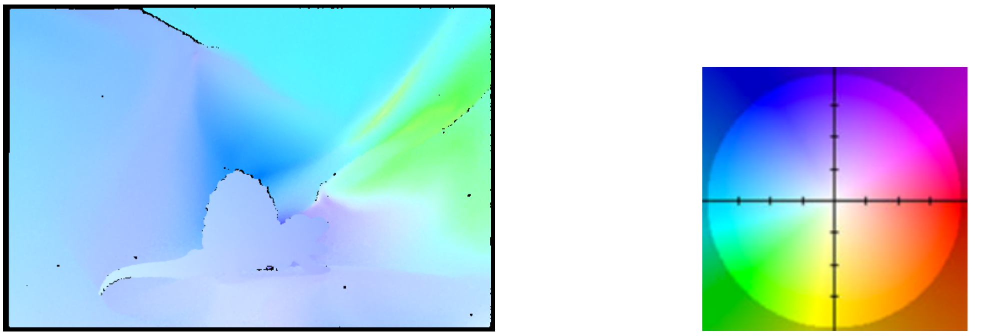


(e) Horn-Schunck (f) Block Matching

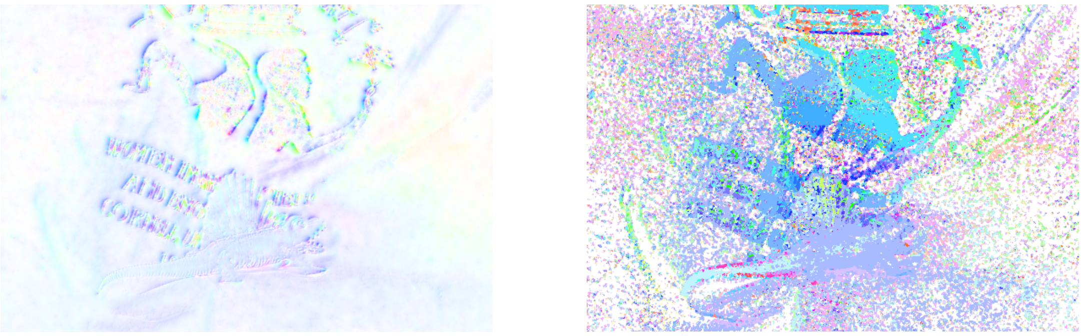
(g) Farneback (h) DeepFlow

Figure 2.13 The input frames of RubberWhale, GT and the estimated flow fields.

(a) Frame 10 (b) Frame 11



(c) GT (d) Colour coding



(e) Horn-Schunck (f) Block Matching

(g) Farneback (h) DeepFlow

Figure 2.14 The input frames of Dimetrodon, GT and the estimated flow fields.

The measurements on Middlebury give consistent results with synthetic images: DeepFlow outperforms the other methods. The accuracy is largely benefited by using the robust SIFT descriptors in feature matching and structuring the response pyramids using multiple sizes whereas the remaining approaches are global algorithms which lack propagation of flow estimation across different scales. The main drawback of DeepFlow is detecting the motion of small objects, as can be observed in Figure 2.14 (h), the shape of the dinosaur's head is very blurred. However, the remainder of the result by DeepFlow is remarkably consistent with the GT, and it is much more evident than for the other (standard) approaches.

2.7 Conclusions

This chapter gives a review of gait analysis and optical flow and evaluates the performance of the benchmark optical flow techniques on both synthetic and real image sequences. The results show that Horn-Schunck cannot handle the large motion. The block size is critical for the performance of Block Matching: if the block size is too small it might not able to integrate all the area, large it will lead the window include the motion belong to other group [48]. The Farneback method is not good at preserving the motion boundaries. DeepFlow appears to outperform the other techniques in most scenarios. The next chapter will introduce how we disambiguate acceleration from optical flow.

Chapter 3 Analysing Acceleration in Computer Images Stream

3.1 Introduction

In most previous research on motion analyses in Computer Vision, only relative movement between consecutive frames has been considered, they did not consider the higher orders of motion. Acceleration is a more distinct feature than displacement, or velocity. Analysing motion in terms of acceleration can provide better understanding of the scene. This chapter introduces our acceleration algorithm and generalize it to real-world motion. In addition the acceleration field is decomposed into constituent parts to allow greater depth of the understanding of the motion and the algorithms have been tested on a variety of image sequences. The experimental results illustrate the ability of acceleration on discriminating different motion whereas velocity did not show any obvious difference.

This chapter is arranged as follows: Section 3.2 introduces and analyses the previous research of acceleration estimation from optical flow. Our new algorithm is introduced in Section 3.3, including the experiments on synthetic and real image sequences. The results show that the new constrains are too strong for real-world motion hence we provide a practical approach in Section 3.4. It outperforms both the previous acceleration algorithms and the new analytic solution. Section 3.5 describes how the resultant acceleration can be decomposed into radial and tangential components, the experimental results demonstrate the ability of acceleration and its constituent parts to disambiguate different types of motion. Finally, the chapter is summarized in Section 3.6.

3.2 Previous Acceleration Algorithms

Acceleration is a vector describing the magnitude and direction of the change of velocity. Average acceleration is the average rate of change of velocity with respect to a time interval. As with velocity, when the time period approaches zero, it is termed instantaneous acceleration:

$$\vec{a} = \lim_{\Delta t \to 0} \frac{\Delta \vec{v}}{\Delta t} = \frac{d\vec{v}}{dt} \qquad (3.1)$$

There was little work analysing acceleration before the work determining gait events through acceleration flow [50]. Since then, there has been contemporaneous research on acceleration which has, as will be shown, significant scope for improvement. Chen and et al. [51] establish an algorithm based on the combination of classic Horn-Schunck [29] and Lucas-Kanade [38] optical flow algorithms. They assume that the image brightness is constant during a short period. Letting $\mathbf{I}(x, y, t)$ denote the image intensity of point (x, y) at time t, then:

$$\mathbf{I}(t-\Delta t)_{x-\Delta x_1, y-\Delta y_1} = \mathbf{I}(t)_{x,y} = \mathbf{I}(t+\Delta t)_{x+\Delta x_2, y+\Delta y_2} \tag{3.2}$$

Expanding the left side of Equation (3.2) by Taylor series gives us:

$$\begin{aligned}\mathbf{I}(t-\Delta t)_{x-\Delta x_1, y-\Delta y_1} &\\ = \mathbf{I}(t)_{x,y} - \mathbf{I}_x\Delta x_1 - \mathbf{I}_y\Delta y_1 - \mathbf{I}_t\Delta t + \frac{1}{2}(\mathbf{I}_{xx}(\Delta x_1)^2) + \frac{1}{2}\left(\mathbf{I}_{yy}(\Delta y_1)^2\right) &\\ + \frac{1}{2}(\mathbf{I}_{tt}(\Delta t)^2) + \mathbf{I}_{xy}\Delta x_1\Delta y_1 + \mathbf{I}_{xt}\Delta x_1\Delta t + \mathbf{I}_{yt}\Delta y_1\Delta t + \xi &\end{aligned} \tag{3.3}$$

where Δx_1 and Δy_1 are the horizontal and vertical displacement between the first frame and second frame; $\mathbf{I}_x = \partial\mathbf{I}/\partial x$, $\mathbf{I}_y = \partial\mathbf{I}/\partial y$, and $\mathbf{I}_t = \partial\mathbf{I}/\partial t$ are the first-order spatial and temporal partial differentials of the image intensity; $\mathbf{I}_{xx}$, $\mathbf{I}_{xy}$, $\mathbf{I}_{yy}$, $\mathbf{I}_{xt}$, $\mathbf{I}_{yt}$ and $\mathbf{I}_{tt}$ are the second-order partial differentials ($\mathbf{I}_{xx} = \partial^2\mathbf{I}/\partial x^2$, etc.), and ξ contains the higher order terms.

The right side of Equation (3.2) is expanded into a similar format as Equation (3.3), after some rearrangement gives:

$$\begin{aligned}\mathbf{I}_{tt}(\Delta t)^2 + \mathbf{I}_{xt}(\Delta x_1 + \Delta x_2)\Delta t + \mathbf{I}_{yt}(\Delta y_1 + \Delta y_2)\Delta t + \mathbf{I}_{xy}(\Delta x_1\Delta y_1 + \Delta x_2\Delta y_2) &\\ + \frac{1}{2}\left(\mathbf{I}_{xx}((\Delta x_1)^2 + (\Delta x_2)^2)\right) + \frac{1}{2}\left(\mathbf{I}_{yy}((\Delta y_1)^2 + (\Delta y_2)^2)\right) &\\ + \mathbf{I}_x(\Delta x_2 - \Delta x_1) + \mathbf{I}_y(\Delta y_2 - \Delta y_1) + \xi = 0 &\end{aligned} \tag{3.4}$$

Equation (3.4) is then divided by $(\Delta t)^2$, meanwhile the higher order term ξ is ignored, resulting in:

$$\begin{aligned}\mathbf{I}_{tt} + \mathbf{I}_{xt}\left(\frac{\Delta x_2}{\Delta t} + \frac{\Delta x_1}{\Delta t}\right) + \mathbf{I}_{yt}\left(\frac{\Delta y_2}{\Delta t} + \frac{\Delta y_1}{\Delta t}\right) + \mathbf{I}_{xy}\left(\frac{\Delta x_1}{\Delta t}\frac{\Delta y_1}{\Delta t} + \frac{\Delta x_2}{\Delta t}\frac{\Delta y_2}{\Delta t}\right) &\\ + \frac{1}{2}\left(\mathbf{I}_{xx}\left(\left(\frac{\Delta x_1}{\Delta t}\right)^2 + \left(\frac{\Delta x_2}{\Delta t}\right)^2\right)\right) + \frac{1}{2}\left(\mathbf{I}_{yy}\left(\left(\frac{\Delta y_1}{\Delta t}\right)^2 + \left(\frac{\Delta y_2}{\Delta t}\right)^2\right)\right) &\\ + \mathbf{I}_x\frac{1}{\Delta t}\left(\frac{\Delta x_2}{\Delta t} - \frac{\Delta x_1}{\Delta t}\right) + \mathbf{I}_y\frac{1}{\Delta t}\left(\frac{\Delta y_2}{\Delta t} - \frac{\Delta y_1}{\Delta t}\right) = 0 &\end{aligned} \tag{3.5}$$

If $\delta t \to 0$, $\frac{\Delta x_1}{\Delta t}$, $\frac{\Delta x_2}{\Delta t}$, $\frac{\Delta y_1}{\Delta t}$ and $\frac{\Delta y_2}{\Delta t}$, are the horizontal and vertical velocities, which are denoted as $u1$, $u2$, $v1$ and $v2$. In their formulation, the terms $\frac{1}{\Delta t}\left(\frac{\Delta x_2}{\Delta t} - \frac{\Delta x_1}{\Delta t}\right)$ and $\frac{1}{\Delta t}\left(\frac{\Delta y_2}{\Delta t} - \frac{\Delta y_1}{\Delta t}\right)$ are considered as the acceleration along x and y axes separately, denoted a_u and a_v:

$$\begin{aligned}\mathbf{I}_{tt} + \mathbf{I}_{xt}(u1+u2) + \mathbf{I}_{yt}(v1+v2) + \mathbf{I}_{xy}(u1v1+u2v2) + \frac{1}{2}\left(\mathbf{I}_{xx}(u1^2+u2^2)\right) \\ + \frac{1}{2}\left(\mathbf{I}_{yy}(v1^2+v2^2)\right) + \mathbf{I}_x a_u + \mathbf{I}_y a_v = 0\end{aligned} \tag{3.6}$$

They assume the velocity is constant, namely $u1 = u2$, $v1 = v2$:

$$\mathbf{I}_{tt} + 2\mathbf{I}_{xt}u + 2\mathbf{I}_{yt}v + 2\mathbf{I}_{xy}uv + \mathbf{I}_{xx}u^2 + \mathbf{I}_{yy}v^2 + \mathbf{I}_x a_u + \mathbf{I}_y a_v = 0 \tag{3.7}$$

Eventually, their constraint equation is:

$$\begin{cases} \mathbf{I}_x a_x + \mathbf{I}_y a_y + \mathbf{I}_v = 0 \\ \mathbf{I}_v = \mathbf{I}_{tt} + 2\mathbf{I}_{xt}u + 2\mathbf{I}_{yt}v + 2\mathbf{I}_{xy}uv + \mathbf{I}_{xx}u^2 + \mathbf{I}_{yy}v^2 \end{cases} \tag{3.8}$$

By following Lucas-Kanande [38], they made an assumption that acceleration is constant over a small patch on the flow field to turn the constraint equations into an over-determined problem. The solution is determined by the least square algorithm.

However, it seems that $\frac{1}{\Delta x}\left(\frac{\Delta x_2}{\Delta t} - \frac{\Delta x_1}{\Delta t}\right)$ and $\frac{1}{\Delta t}\left(\frac{\Delta y_2}{\Delta t} - \frac{\Delta y_1}{\Delta t}\right)$ in Equation (3.5) probably indicate the derivatives of acceleration: Jerk. If our analysis is true, their algorithm derives jerk fields rather than acceleration fields. We were unfortunately unable to verify this with the author.

Later, Dong and et al. [52] aimed to feed acceleration into deep networks as a motion descriptor for detecting violence in videos. In their work, acceleration was estimated by expanding the horizontal velocity field $\mathbf{U}$ and vertical velocity $\mathbf{V}$ at $t + \Delta t$ by a Taylor series:

$$\begin{aligned}\mathbf{U}(t+\delta t)_{x+\Delta x, y+\Delta y} &= \mathbf{U}(t)_{x,y} + \mathbf{U}_x \Delta x + \mathbf{U}_y \Delta y + \mathbf{U}_t \Delta t \\ \mathbf{V}(t+\delta t)_{x+\Delta x, y+\Delta y} &= \mathbf{V}(t)_{x,y} + \mathbf{V}_x \Delta x + \mathbf{V}_y \Delta y + \mathbf{V}_t \Delta t\end{aligned} \tag{3.9}$$

the higher order terms are ignored. The change of velocity over time is:

$$\begin{aligned}\frac{d\mathbf{U}}{dt} &= \mathbf{U}_x \frac{dx}{dt} + \mathbf{U}_y \frac{dy}{dt} \\ \frac{d\mathbf{V}}{dt} &= \mathbf{V}_x \frac{dx}{dt} + \mathbf{V}_y \frac{dy}{dt}\end{aligned} \tag{3.10}$$

The approach appeared to assume that $\left(\frac{dx}{dt}, \frac{dy}{dt}\right)$ corresponds to (u, v), therefore, they obtain the acceleration as:

$$\mathbf{a} = (\omega \Delta u, \omega \Delta v) \tag{3.11}$$

They obtain the acceleration flow by a second-order differential of neighbouring frames, which is computing flow on optical flow. However, the main drawback of this idea probably is it is difficult to compute the spatial partial derivatives due to the smoothness of the optical flow field (the neighbour pixels tend to have similar velocities), this is also the reason why we abandon this method at the beginning of our research.

There appear to be errors in Chen's solution [51] and perhaps the main weakness of their algorithm is that the constraints are too strong. In Dong's algorithm [52], it is hard to compute the optical flow on optical flow due to the smoothness of flow field. Furthermore, there is no quantitative experimental analysis of the algorithms in their papers. Therefore, they remain an unsolved problem and it worthy of further investigation since acceleration can be used to disambiguate motion. In the next section, we will describe our solution to this problem.

3.3 Estimation of Acceleration Flow

3.3.1 Recovering Acceleration from Optical Flow

The accuracy of optical flow algorithms has improved steadily over the past few years, the basic formulation, changed little after the pioneering work of Horn and Schunck [47]. We build our variational acceleration algorithm based on their work since almost all the state of the art algorithms are still using their theory [48].

They assumed that the intensity of a point does not change between two consecutive frames. Now we extend this principle to three frames for estimating acceleration fields. If $\mathbf{I}(x, y, t)$ denotes the image intensity on (x, y) at time t, the image intensity is constant at frame $t - \delta t$, t, $t + \delta t$:

$$\begin{cases} \mathbf{I}(t-\Delta t)_{x-\Delta x_1, y-\Delta y_1} = \mathbf{I}(t)_{x,y} \\ \mathbf{I}(t)_{x,y} = \mathbf{I}(t+\Delta t)_{x+\Delta x_2, y+\Delta y_2} \end{cases} \tag{3.12}$$

Expanding Equation (3.12) by a Taylor expansion:

$$\begin{cases} \mathbf{I}(t)_{x,y} - \mathbf{I}_x \Delta x_1 - \mathbf{I}_y \Delta y_1 - \mathbf{I}_t \Delta t + \xi = \mathbf{I}(t)_{x,y} \\ \mathbf{I}(t)_{x,y} = \mathbf{I}(t)_{x,y} + \mathbf{I}_x \Delta x_2 + \mathbf{I}_y \Delta y_2 + \mathbf{I}_t \Delta t + \xi \end{cases} \tag{3.13}$$

and ignoring the higher order terms, we have:

$$-\mathbf{I}_x \Delta x_1 - \mathbf{I}_y \Delta y_1 - \mathbf{I}_t \Delta t = \mathbf{I}_x \Delta x_2 + \mathbf{I}_y \Delta y_2 + \mathbf{I}_t \Delta t \tag{3.14}$$

Dividing Equation (3.14) by Δt,

$$-\mathbf{I}_x \frac{\Delta x_1}{\Delta t} - \mathbf{I}_y \frac{\Delta y_1}{\Delta t} - \mathbf{I}_t = \mathbf{I}_x \frac{\Delta x_2}{\Delta t} + \mathbf{I}_y \frac{\Delta y_2}{\Delta t} + \mathbf{I}_t \tag{3.15}$$

then the gradient constraint is yielded:

$$\nabla\mathbf{I}\cdot(\mathbf{v}_{t-\Delta t})-\mathbf{I}_t=\nabla\mathbf{I}\cdot(\mathbf{v}_{t+\Delta t})+\mathbf{I}_t \tag{3.16}$$

where $\nabla=\left(\frac{\partial}{\partial x},\frac{\partial}{\partial y}\right)$, and $\mathbf{v}$ consists of horizontal and vertical components $(u,v)^T$.

If the acceleration changes dynamically from frame to frame, then there is little chance of recovering the acceleration. More commonly, motion is smooth which means acceleration is constant during a small period. Therefore, we assume that the acceleration does not change during three consecutive frames. According to Newton's laws:

$$\mathbf{v}_t=\mathbf{v}_0+\mathbf{a}\Delta t \tag{3.17}$$

$\mathbf{v}_{t-\Delta t}$ and $\mathbf{v}_{t+\Delta t}$ in Equation (3.16) can be substituted by:

$$\begin{cases}\mathbf{v}_{t-\Delta t}=\mathbf{v}_t-\mathbf{a}\Delta t\\ \mathbf{v}_{t+\Delta t}=\mathbf{v}_t+\mathbf{a}\Delta t\end{cases} \tag{3.18}$$

where $\mathbf{v}_t$ represents the velocity vector at time t and the acceleration vector $\mathbf{a}$ is composed of horizontal and vertical components $(a_u,a_v)^T$, we have:

$$\nabla\mathbf{I}(\mathbf{v}_t-\mathbf{a}\Delta t)-\mathbf{I}_t=\nabla\mathbf{I}(\mathbf{v}_t+\mathbf{a}\Delta t)+\mathbf{I}_t \tag{3.19}$$

$$\nabla\mathbf{I}\cdot\mathbf{a}\Delta t+\mathbf{I}_t=0 \tag{3.20}$$

Dividing Equation (3.20) by Δt:

$$\nabla\mathbf{I}\cdot\mathbf{a}+\frac{\mathbf{I}_t}{\Delta t}=0 \tag{3.21}$$

If $\Delta t\to 0$, our optical flow constraint equation of acceleration is:

$$\nabla\mathbf{I}\cdot\mathbf{a}+\mathbf{I}_{tt}=0 \tag{3.22}$$

where $\mathbf{I}_{tt}$ is the second order of image intensity with respect to time. Now we have two unknowns $\mathbf{a}(a_u,a_v)^T$ in Equation (3.22) but with only one equation, there are infinite solutions. Hence, we need seek another equation to avoid this ill-posed problem.

Acceleration also has similar smoothness characteristics to velocity in that neighbouring pixels tend to have similar acceleration. This shows a natural linkage between velocity and acceleration analysis in image sequences. The smoothness means that the neighbouring pixels tend to have similar acceleration. The way to estimate the smoothness constrain is by minimizing the error of the squares of the Laplacians of the horizontal and vertical flow:

$$\varepsilon_s^2=\iint(\nabla^2a_u+\nabla^2a_v)\,dxdy=\iint\left(\frac{\partial^2a_u}{\partial x^2}+\frac{\partial^2a_u}{\partial y^2}+\frac{\partial^2a_v}{\partial x^2}+\frac{\partial^2a_v}{\partial y^2}\right)dxdy \tag{3.23}$$

Combing the error needing to be minimized into brightness constrain gives:

$$\varepsilon^2 = \iint \left(\varepsilon_d^2 + \lambda^2 \varepsilon_s^2\right) dxdy \tag{3.24}$$

where λ is dependent on the noise in the pixel intensity, the error of data term ε_d is:

$$\varepsilon_d = \iint (\nabla \mathbf{I} \cdot \mathbf{a} + \mathbf{I}_{tt})\, dxdy \tag{3.25}$$

Now we have the problems well-posed, we will describe the implementation of our algorithm in the next section.

3.3.2 Approximating the Derivatives

For estimating acceleration in image sequences, the derivatives are approximated between three consecutive frames. It is important that the derivatives are consistent so we refer to the same point in the image at same time in the implementation. There are many ways of estimating differentiation, we use the same spatial kernels with Horn-Schunck [29]. The spatial-temporal relationship is illustrated in Figure 3.1.

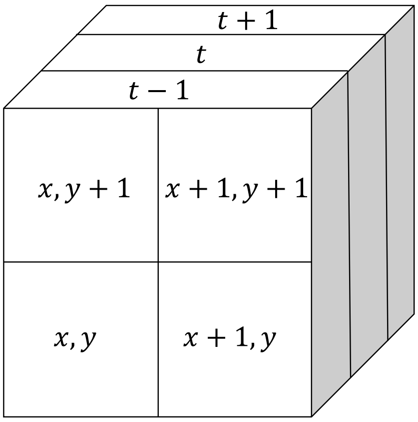


Figure 3.1 Estimating the partial derivatives for point (x, y, t).

For estimating the first order horizontal and vertical derivatives, we use the following kernel [29]:

-1	1
-1	1

1	1
-1	-1

Figure 3.2 Spatial derivatives kernels.

Since we need to consider three frames in computing acceleration the spatial derivatives are:

$$\begin{aligned}\mathbf{I}_x &\approx \frac{1}{6}\sum_{y}\sum_{t}\left(\mathbf{I}_{x+1,y,t} - \mathbf{I}_{x,y,t}\right)\\ \mathbf{I}_y &\approx \frac{1}{6}\sum_{x}\sum_{t}\left(\mathbf{I}_{x,y+1,t} - \mathbf{I}_{x,y,t}\right)\end{aligned} \tag{3.26}$$

The data constrain (Equation (3.22)) contains the second-order time derivative, we choose Laplacian operator to compute the second order of time derivative in the implementation, the template is:

1	-2	1

Figure 3.3 The temporal temple for second order time derivative.

$$\mathbf{I}_{tt} \approx \frac{1}{8}\sum_{x}\sum_{y}\left(\mathbf{I}_{x,y,t+1} - 2\mathbf{I}_{x,y,t} + \mathbf{I}_{x,y,t-1}\right) \tag{3.27}$$

where $x \in \{m, m+1\}, y \in \{n, n+1\}, t \in \{k-1, k, k+1\}$.

By following a similar solution to Horn-Schunck to express the smoothness constraint, eventually we can determine the acceleration flow in image by:

$$\begin{aligned}\left(\mathbf{I}_x^2 + \mathbf{I}_y^2\right)\left(a_u - \overline{a_u}\right) &= -\mathbf{I}_x\left[\mathbf{I}_x\overline{a_u} + \mathbf{I}_y\overline{a_v} + \mathbf{I}_{tt}\right]\\ \left(\mathbf{I}_x^2 + \mathbf{I}_y^2\right)\left(a_v - \overline{a_v}\right) &= -\mathbf{I}_y\left[\mathbf{I}_x\overline{a_u} + \mathbf{I}_y\overline{a_v} + \mathbf{I}_{tt}\right]\end{aligned} \tag{3.28}$$

We now have the basis for detecting acceleration, we shall now move to evaluating this algorithm to determine whether we can indeed detect acceleration from image intensity.

3.3.3 Analysing Acceleration Algorithm on Image Sequences

We evaluate Variational-Acceleration first on synthetic images to assess performance before analysis on real images to assess its capabilities. Figure 3.4 shows the acceleration detection results of synthetic image sequences with rigid motion. The algorithms are tested on the sequences without acceleration, with small acceleration and large acceleration respectively. The results detected by our Variational-Acceleration algorithm are compared with Dong's algorithm (denote by Dong) [52]. Dong detects acceleration flow by computing the optical flow on optical flow, but they did not specify which method they used in their paper. To make it fair, we use Horn-Schunck in the implementation of their method since our acceleration algorithm is based on Horn-Schunck as well.

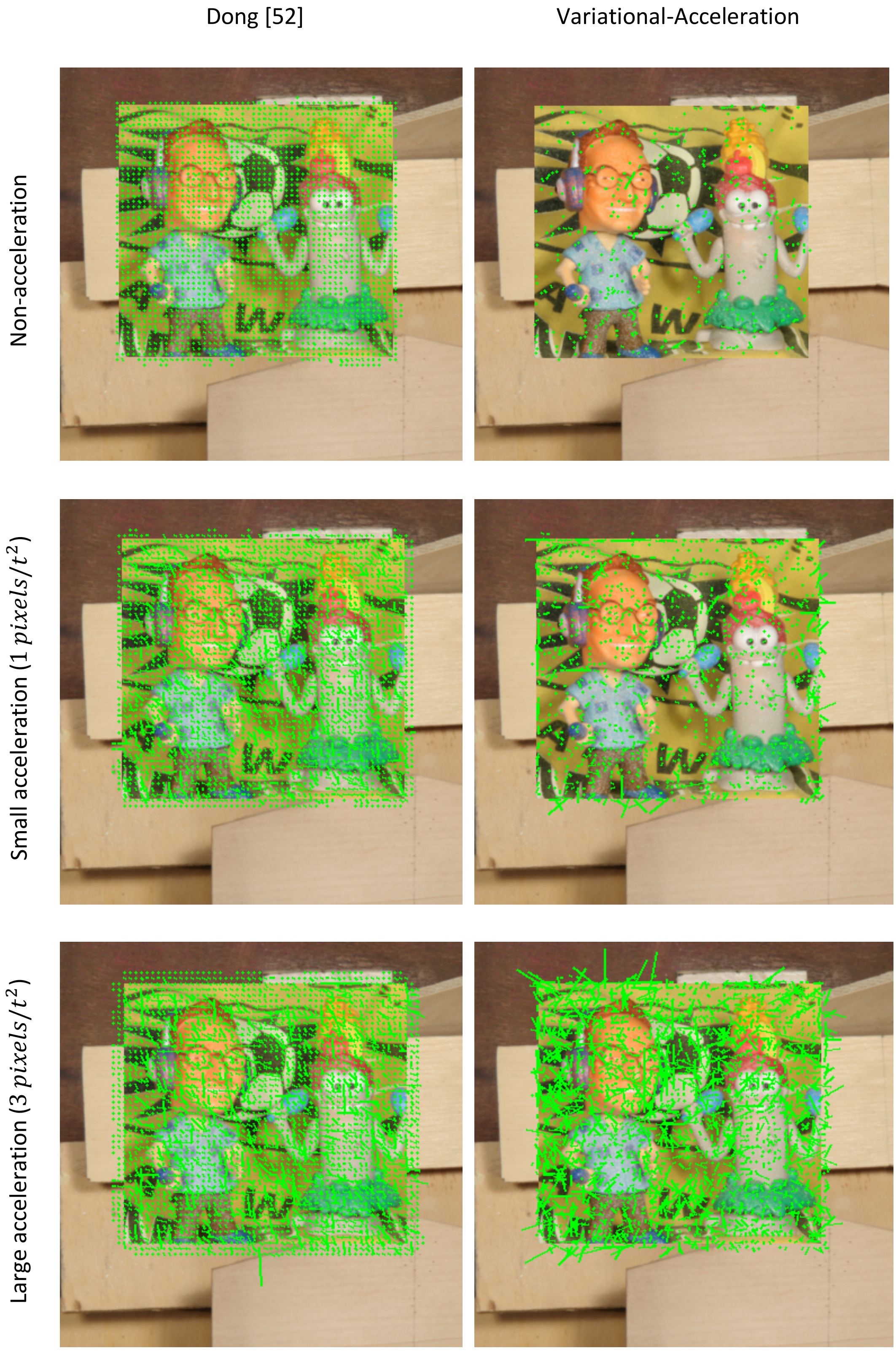


Figure 3.4 The acceleration fields detected by various methods of synthetic images.

In the first row, Dong detects evenly distributed acceleration field on the moving block however the motion field does not contain any acceleration. The acceleration field become denser when the sub-Mequon is accelerating with small value in the second row and it did not change much when the acceleration become large.

In the performance of Variational-Acceleration, the result is acceptable when the sub-Mequon moves constantly. There is little acceleration flow detected, this is consistent with the expected motion, only some random noise. Variational-Acceleration detects more flow with the increasing of acceleration but the results tending to be noisier at the same time. The results of both methods are not satisfactory, we seek another method to obtain the optimized results of acceleration field in the next section.

3.4 Estimating Acceleration Flow via Other Flow Estimation Methods

3.4.1 A More Practical Approach

Since the motion in real images is often large and non-rigid, we want to seek more feasible form for recovering acceleration from image sequences. Optical flow is still an active area in computer vision, and new algorithms are constantly emerging and the performance has significantly improved since the first variational algorithm Horn-Schunck. Therefore, instead of aiming to improve the basic theory, we think it was better to use a state of art algorithm to approximate the acceleration flow. According to Equation (3.1), the acceleration field can be approximated by the differential of neighbour velocity fields:

$$\widehat{\mathbf{A}}(t) = \widehat{\mathbf{V}}\big(t{\sim}(t+\Delta t)\big) - \widehat{\mathbf{V}}\big((t-\Delta t){\sim}t\big) \qquad (3.29)$$

where $\widehat{\mathrm{V}}\big(t{\sim}(t+\Delta t)\big)$ is the velocity field referred frame t to $(t+\Delta t)$ and $\widehat{\mathrm{V}}\big((t-\Delta t){\sim}t\big)$ is the velocity field referred frame $t-\Delta t$ to t. In implementation, the resultant optical flow field can be considered as the velocity field due to the fix frame rate, the unit is pixel/frame.

To avoiding the error caused by inconsistent reference along the time axis, we proposed a new way that refers to the middle frame as the start frame in the temporal template as explained in Figure 3.5, we term this approach Differential-Acceleration algorithm. The acceleration field approximation by Differential-Acceleration can be expressed as:

$$\widehat{\mathbf{A}}(t) = \widehat{\mathbf{V}}\big(t{\sim}(t+\Delta t)\big) - \big\{-\widehat{\mathbf{V}}\big(t{\sim}(t-\Delta t)\big)\big\} \qquad (3.30)$$

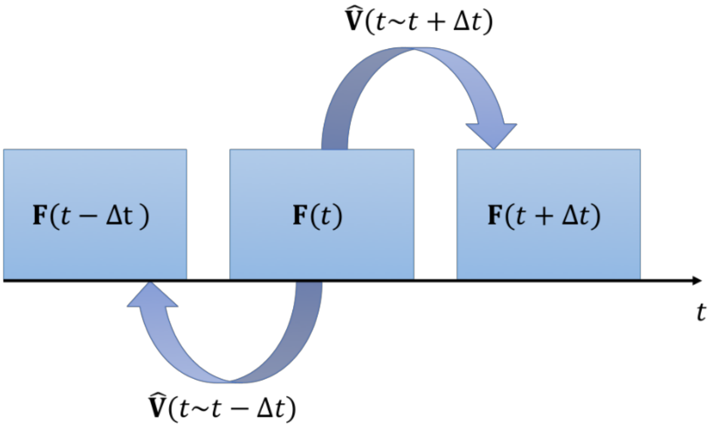


Figure 3.5 Computing acceleration field referred the middle frame as the start.

Now we are going to deploy this more practical algorithm on images to see whether this approach can achieve better results than Variational-Acceleration. In this thesis, DeepFlow [36], which is a popular new technique with excellent performance for large displacement estimation and non-rigid matching, is used as our fundamental technique for Differential-Acceleration.

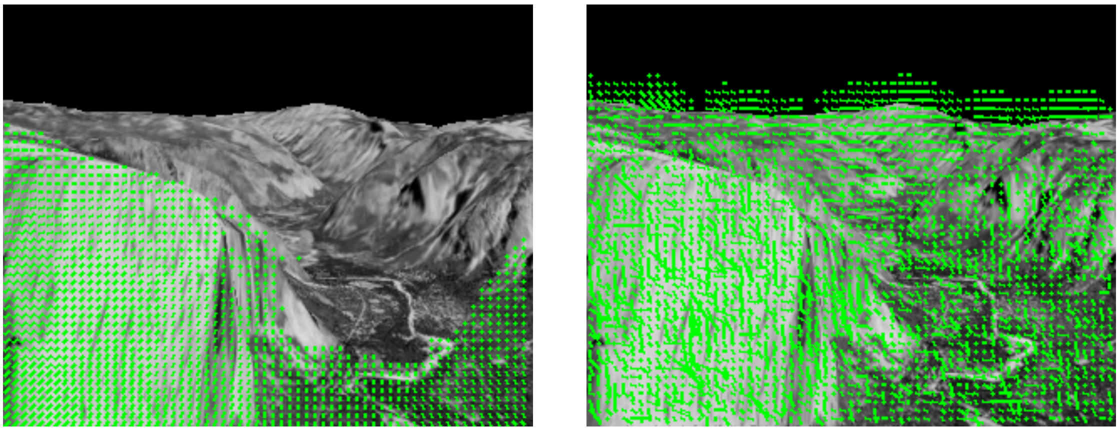
3.4.2 Evaluating Acceleration Algorithms on Synthetic Images

The performances of Dong [52], Variational-Acceleration and Differential-Acceleration are evaluated on the famous test sequence Yosemite. There is little acceleration in the original image sequence, so we induce artificial acceleration in this image sequence by skipping one frame to see whether the algorithms are robust enough to estimate acceleration fields under this sequence. The detection results are illustrated in Figure 3.6.

Since currently most optical flow databases [7], [53] are targeted at measuring velocity flow, either the test image sequences only have two frames or the GT is only one single velocity field. Obtaining an accurate GT of these benchmark optical flow databases became an obstacle during this research on acceleration. For evaluating the acceleration algorithms, pseudo acceleration GT is computed by MDP-Flow2 [34], it is a highly-ranked method on the optical flow algorithms evaluation[4] and the code is publicly available. We use MDP-Flow2 to estimate the pseudo GT of velocity flow first, the pseudo acceleration flow is computed by Equation (3.30).

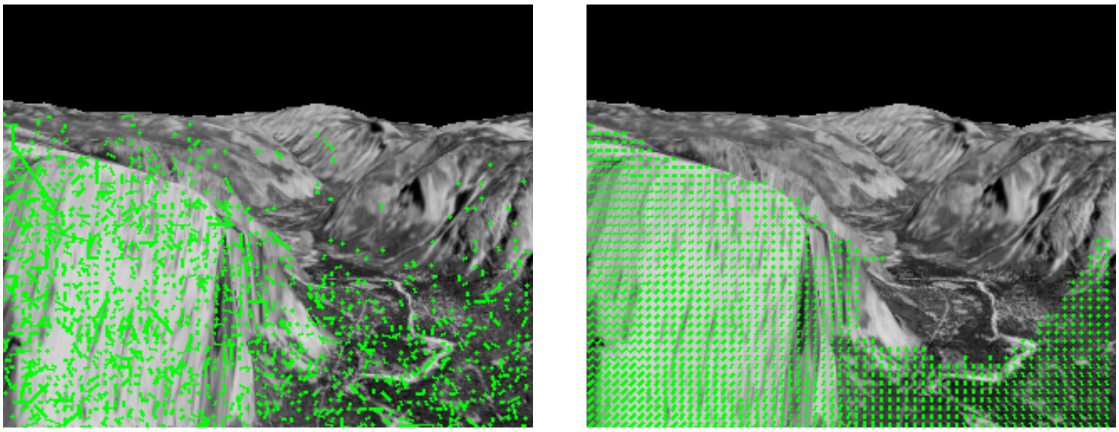
---

[4] http://vision.middlebury.edu/flow/eval/



(a) Pseudo GT estimated by MDP-Flow2 [34]

(b) Dong [52]



(c) Variational-Acceleration.

(d) Differential-Acceleration

Figure 3.6 Comparing detection results on Yosemite.

The acceleration in pseudo GT mostly focusses on the lower part of the manipulated Yosemite sequence shown in Figure 3.6 (a). The detection results of Dong in (b) show the flow is on the whole mountain area. This is because the principle of their method computes the optical flow of the optical flow. However, once the flow is detected from the image sequence, the field varies in a smooth manner, that means the neighbour pixels tend to have similar velocities, the moving objects lose texture which is important for the majority of optical flow algorithms.

In (c) the results estimated by Variational-Acceleration show an improvement over Dong, with the acceleration mostly focussed on the lower left corner in the image which shows consistency with the pseudo GT. However, the result is still very noisy because the motion violates both the data and smoothness constrains (occlusion and large motion). Variational-Acceleration restricts to the linear domain of the image derivatives, which means the displacement has to be very small and the motion must be smooth.

Encouragingly, Differential-Acceleration detected evenly distributed acceleration flow in the lower right corner. The result shows a considerable improvement compared with the results of Dong and Variational-Acceleration. The acceleration field in (d) shows less noise and it is more the consistent with the pseudo GT.

3.4.3 Comparison between Differential and Variational-Acceleration

Apart from Yosemite, Table 3.1 and Table 3.2 reports the statistics of performance of Dong, Variational-Acceleration and Differential-Acceleration in estimating acceleration on a number of sequences of the Middlebury [7] for a more objective evaluation. The statistical results are compared with the pseudo GT. The pseudo GT is estimated by MDP-Flow2 [34] since only one ground truth optical flow field between consecutive frames is available on the evaluation website for preventing upcoming new algorithms to over fit the test images. AEPE is chosen here for presenting the error since the motion in Middlebury image sequences are relatively large in the pseudo GT. The average ranks, reported AEPE and SD of MDP-Flow2, DeepFlow and Horn-Schunck on the Middlebury evaluation website are shown in Table 3.3 and Table 3.4 for reference.

Method	Backyard	Dumptruck	Mequon	Schefflera	Walking	Yosemite	Avg.
Dong [52]	3.21	2.58	3.54	3.35	3.13	3.39	3.2
Variational-Acceleration	2.48	1	3.38	2.89	1.87	2.1	2.29
Differential-Acceleration	0.35	0.3	0.29	0.37	0.51	0.25	0.35

Table 3.1 AEPE of estimation algorithms on Middlebury datasets.

Method	Backyard	Dumptruck	Mequon	Schefflera	Walking	Yosemite	Avg.
Dong [52]	2.94	2.43	3.01	2.67	2.45	3.26	2.79
Variational-Acceleration	3.74	2.79	4.18	3.49	2.52	2.62	3.22
Differential-Acceleration	0.94	1.29	0.78	0.98	0.99	0.32	0.88

Table 3.2 SD of EPE between acceleration algorithms.

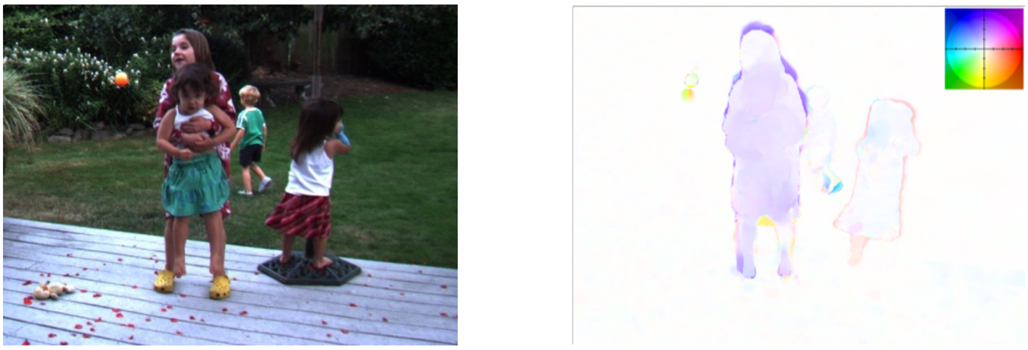
Method	AEPE Average Rank	Mequon	Schefflera	Yosemite
MDP-Flow2 [34]	11.8	0.15	0.20	0.11
DeepFlow [36]	69.9	0.28	0.44	0.11
Horn-Schunck [29]	113.3	0.61	1.01	0.16

Table 3.3 The reported AEPE ranks on Middlebury evaluation website[5].

Method	Mequon	Schefflera	Yosemite
MDP-Flow2 [34]	0.40	0.55	0.12
DeepFlow [36]	0.78	1.23	0.12
Horn-Schunck [29]	0.98	1.88	0.16

Table 3.4 The reported SD on Middlebury evaluation website[5].

We follow the convention of optical flow evaluation which uses average absolute endpoint error and the standard deviation to present the results. It is difficult to include the uncertainty although it will give more information of the results since the error of each part could be very different (e.g. some algorithms are good on the discontinuous area and some are good on the textured area). The benchmark evaluation dataset Middlebury gives the separate results of "all", "discontinuous" and "textured". However, it is difficult for us (and other researchers) to separate the results because the labels for the edges and textured area are not available.



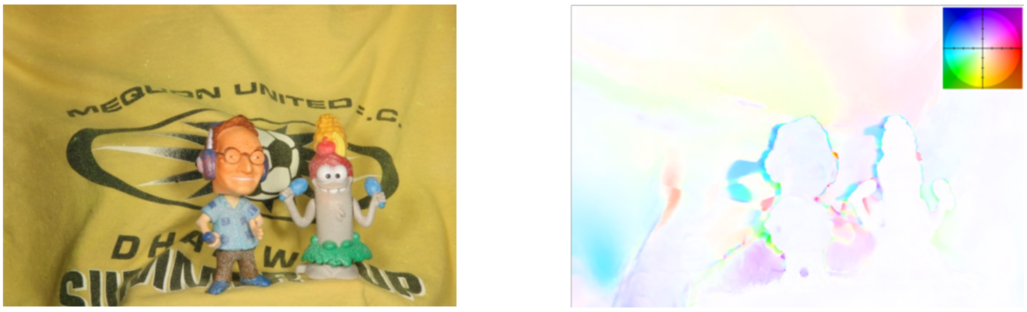
(a) Backyard

(b) Pseudo GT

---

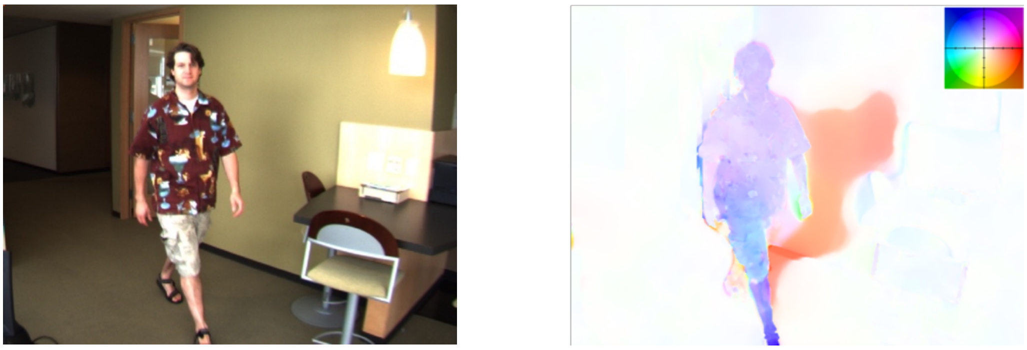
[5] http://vision.middlebury.edu/flow/eval/results/results-a1.php

(c) Variational-Acceleration (d) Differential-Acceleration



(e) Mequon (f) Pseudo GT

(g) Variational-Acceleration (h) Differential-Acceleration



(i) Walking (j) Pseudo GT

(k) Variational-Acceleration (l) Differential-Acceleration

Figure 3.7 More examples of detecting acceleration on real images [7].

Examples of results from Middlebury dataset are illustrated in Figure 3.7. The motion of illustrated results are relatively dense so visualizing motion field with arrows could cause confusion, we opt to use flow field colour coding create by [7] to present the fields. The experimental results show that both the assumptions of constant intensity and smooth motion in Variational-Acceleration are too strong for real motion (most are complex) so we will use Differential-Acceleration to detect acceleration flow in the rest of the thesis.

We now have the basis for detecting acceleration: a Variational Approach, Variational-Acceleration, and its extension to a more generalized Differential Approach form, Differential-Acceleration that is shown to be able to detect acceleration in complex scenes. In the next section, the acceleration is going to be decomposed into radial and tangential components to aid further understanding of complex motion.

3.5 Tangential and Radial Acceleration

3.5.1 Decomposing the Resultant Acceleration

Acceleration of curved motion is composed of two components: tangential and radial acceleration. The tangential component changes the magnitude of the velocity and the direction is located in the tangent line of the trajectory (increasing or decreasing the speed). The radial component (also called centripetal acceleration in circular motion) changes the direction of the velocity and it points to the centre of the curved path (normal to the tangent line of the trajectory), as shown in as shown in Figure 3.8.

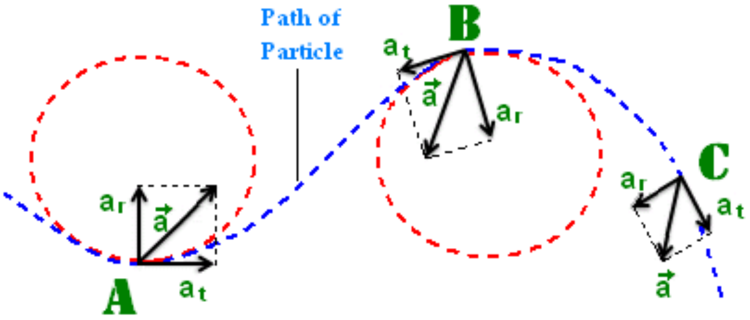


Figure 3.8 The relationship between resultant acceleration, tangential acceleration and radial acceleration[6].

There is linear or circular motion, therefore, the motion incorporated in images is either linear or circular if the time interval is sufficiently small. We assume that the moving points which follow the curved trajectories rotate along the same arc in any three consecutive frames since three nonlinear points can determine one and only one circle. The rotation centre can be calculated by the positions of the pixel in the consecutive frames. Connecting these three points with straight lines and applying perpendicular bisectors to them, the centre of the circle is then located at the intersection of the two perpendicular bisectors, as shown in Figure 3.9.

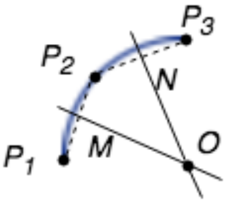


Figure 3.9 Location of radial acceleration centre.

Suppose the coordinates of a point in three consecutive frames are: $P_i(x_i, y_i)\ i \in (1,2,3)$, MO, NO are the perpendicular bisectors of P_1P_2 and P_2P_3, and $O(x_o, y_o)$ denotes the center of the circular motion, hence:

$$\overrightarrow{MO} \cdot \overrightarrow{P_1P_2} = \overrightarrow{NO} \cdot \overrightarrow{P_2P_3} \tag{3.31}$$

Then the coordinate of *O* can be obtained by:

$$O^T = 0.5 \cdot \mathbf{\Phi}^{-1}\mathbf{\Psi} \tag{3.32}$$

where,

---

[6] Image is taken from: http://physics.tutorvista.com/motion/tangential-acceleration.html

$$\mathbf{\Phi} = \begin{bmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_2 & y_3 - y_2 \end{bmatrix} \quad \mathbf{\Psi} = \begin{bmatrix} {x_2}^2 - {x_1}^2 + {y_2}^2 - {y_1}^2 \\ {x_3}^2 - {x_2}^2 + {y_3}^2 - {y_2}^2 \end{bmatrix} \tag{3.33}$$

We use $\mathbf{a} = (x_2 + a_u, y_2 + a_v)$ to present the coordinates of acceleration vector in image plane. The positions of tangential acceleration $\boldsymbol{tang}(u,v)$ and radial acceleration $\boldsymbol{rad}(u,v)$ can be estimated by:

$$\begin{aligned} \boldsymbol{tang}^T &= [\mathbf{f}(-\theta)\ \ \mathbf{g}(-\theta)]^T [\mathbf{p}_2 \cdot \mathbf{f}(\theta) \quad \mathbf{a} \cdot \mathbf{g}(\theta)\ ]^T \\ \boldsymbol{rad}^T &= [\mathbf{f}(-\theta)\ \ \mathbf{g}(-\theta)]^T [\ \mathbf{a} \cdot \mathbf{f}(\theta) \quad \mathbf{p}_2 \cdot \mathbf{g}(\theta)]^T \end{aligned} \tag{3.34}$$

where θ is the angle between OP_2 and the horizontal axis $\mathbf{f}(\theta) = (\cos\theta, \sin\theta)$, $\mathbf{g}(\theta) = (-\sin\theta, \cos\theta)$.

We now have the basis for detecting acceleration and an extension to a more generalized decomposed form. In the next section, we evaluate these approaches to determine whether we can indeed detect acceleration from image sequences.

3.5.2 Deploying New Algorithms on Image Sequences

We evaluate the decomposed approach first on synthetic images to assess performance before analysis on real images to show its capability in real applications. The experiments are classified into 4 groups: linear shift with constant velocity, linear shift with acceleration, rotation in constant angular velocity and rotation with angular acceleration. We detected velocity, resultant acceleration, radial and tangential fields for each group separately and the results are illustrated in Figure 3.10.

Figure 3.10 The experimental results of synthetic images.

In the results of linear shift, there is little radial acceleration because the direction of trajectory does not change, and the tangential component only appears when the object is accelerating. The resultant acceleration field shows similar features with tangential acceleration since it only contains the tangential component in linear displacement. Velocity appears in both situations however.

In the rotation examples, radial acceleration both appears under rotation with constant angular velocity and angular acceleration, due to the direction of motion changing all the time. The magnitude of radial acceleration is increasing with the angle of the object rotated. All the radial acceleration flow points to the sub-Mequon centre since it rotates about the centre. The directions of tangential components are along the tangent of rotating trajectory showing a result consistent with expectations. Simultaneously, the velocity field does not show any obvious distinction. There appears to be some noise around the edge of the moving object; this is mainly caused by discontinuous motion in that area. Our estimated detection of acceleration shows expected results on artificial scenes, they illustrate detection of features consistent with acceleration features, which velocity analysis lacks.

In real image sequences, acceleration flow can help distinguish objects undergoing different motions. Figure 3.11 and Figure 3.12 gives different flow fields of two images sequences [7] which are natural motion in real world by high speed camera. In the first row, the velocity fields contain any types of motion so they look chaos, as contrast the motion in radial and tangential acceleration fields are more comprehensible.

In first sequence “Basketball”, the left man is passing the ball to another person. At the beginning of the throw, the ball has acceleration and it hasn’t started spinning so the acceleration is mainly tangential in frame 9, the left man’s hands shows little tangential acceleration flow since he has not raised his hand to take the pass. In frame 11, the radial acceleration reveals that the basketball is spinning during the travel, as well as the shadow on the wall. The tangential acceleration shown on the hand of the right man indicates that he raises his hands and ready to take the pass. The acceleration fields distinguish the motion field features as known. In the second sequence “Backyard” the little girls are jumping up. The acceleration flow shows they have the most tangential acceleration at the start of the jump and little in the end.

Figure 3.11 Different flow fields from Middlebury dataset Basketball.

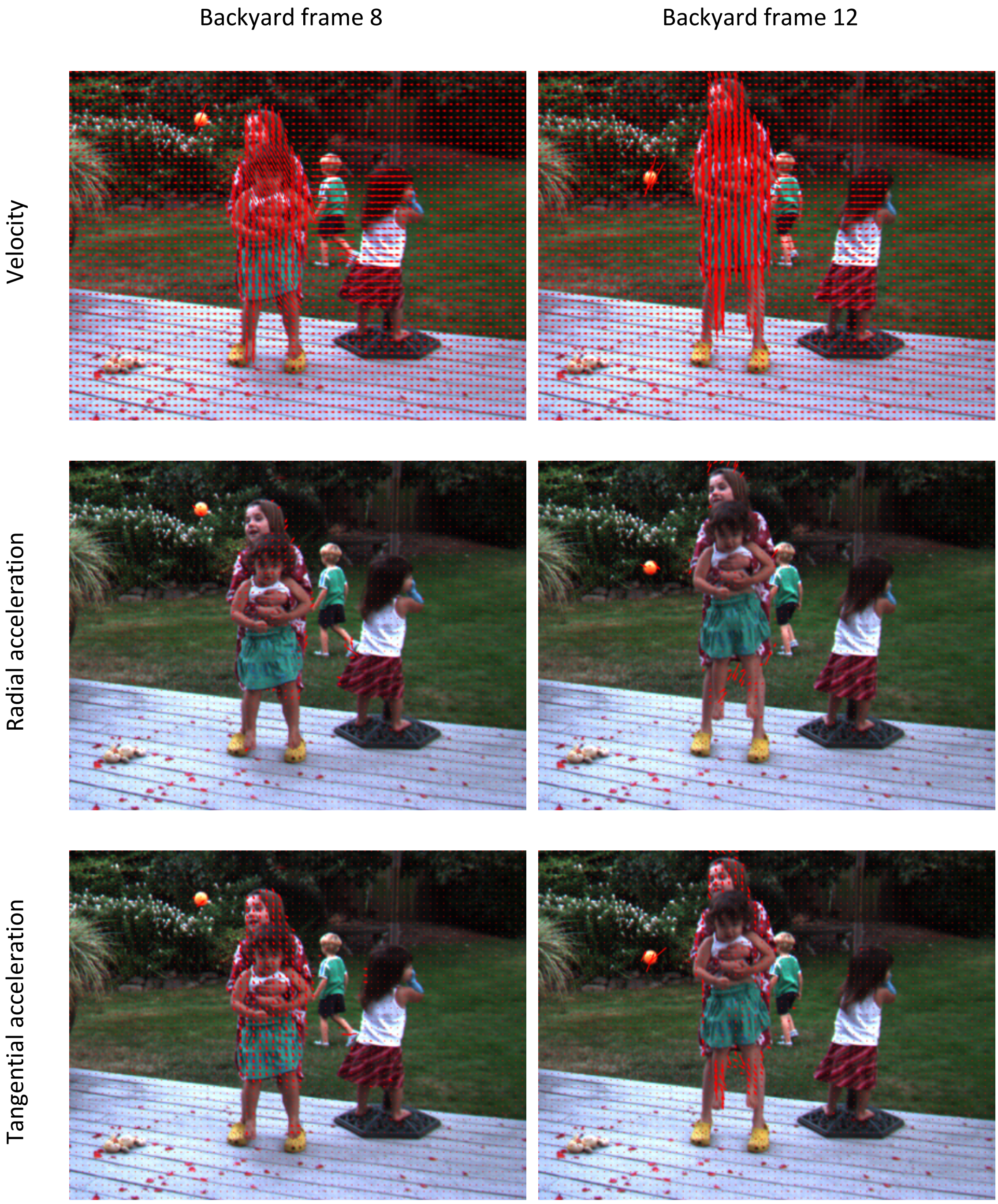


Figure 3.12 Different flow fields from Middlebury dataset Backyard.

In Figure 3.11 a significant amount of radial acceleration appears on the spinning basketball and its shadow on the wall. The accumulated rotation centre in Figure 3.13 (the blue cross on the right-hand man) is consistent with the direction of radial acceleration on the shadow since it has much larger flow filed than the basketball itself. The centres are accumulated by the proximity algorithm [28] and the radius is 10 pixels. The result demonstrates the power of our acceleration decomposition algorithm and the information can benefits the prediction of the trajectory the moving objects.

Figure 3.13 Accumulated radial acceleration rotation centre in test sequence "Basketball".

We also provide a new example of our approach detecting acceleration on gait in a Chroma key laboratory [54], shown in Figure 3.14. Acceleration is detected mainly only around the limbs of the walking subject and is maximum around the swinging forward leg while the other leg is in the stance phase, since the limbs appear to have pendulum-like motion when people are walking [8]. Beyond that Figure 3.15 shows the zoom-in radial acceleration on the subject’s thorax, the direction of radial acceleration reveals that the upper body actually moves like a pendulum to a small extent during walking, which is consistent with the argument in [8]. In contrast, the velocity flow is distributed all over the body without notable difference, so the detected acceleration is consistent with the above analysis.

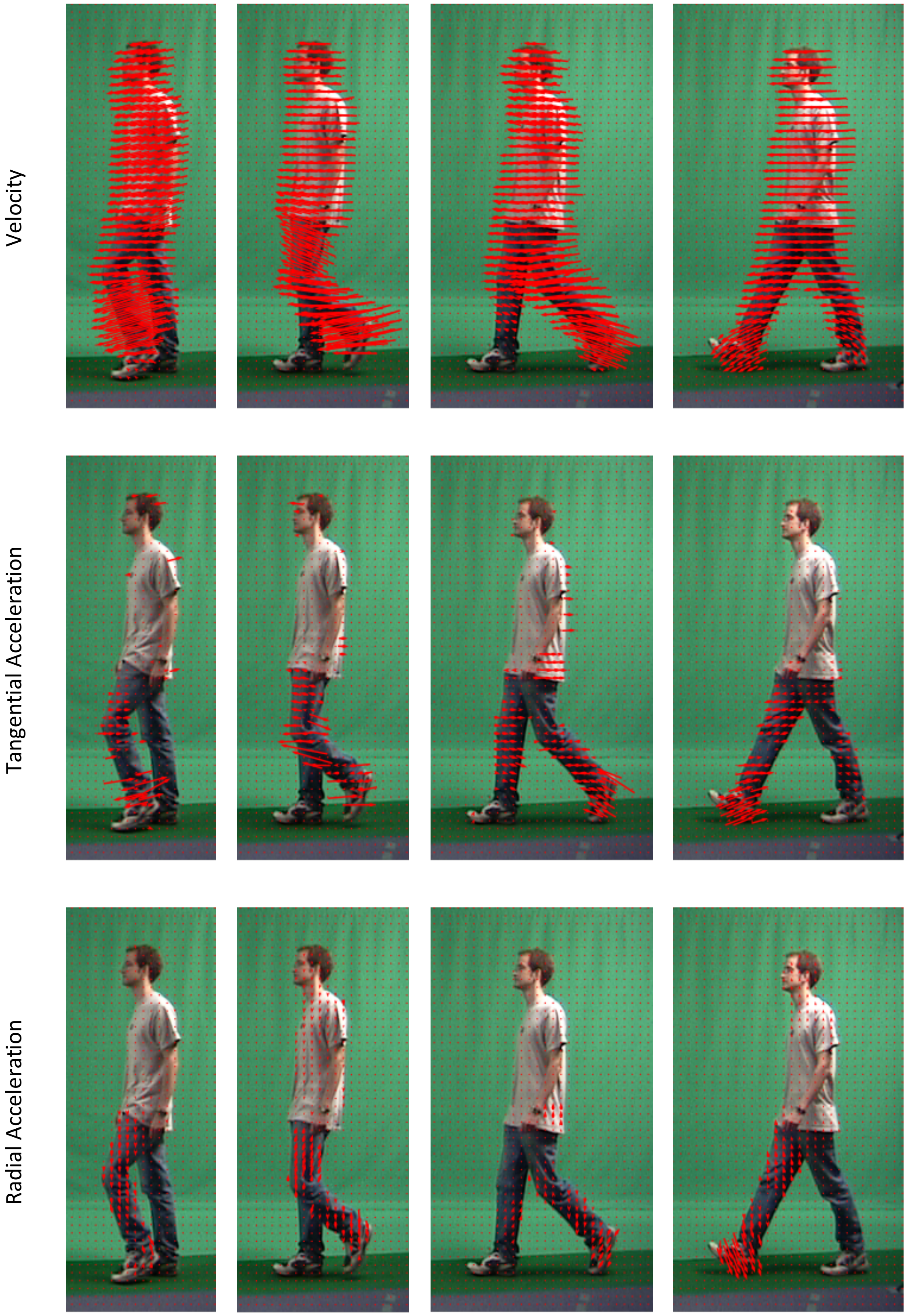


Figure 3.14 The acceleration fields of half walking cycle.

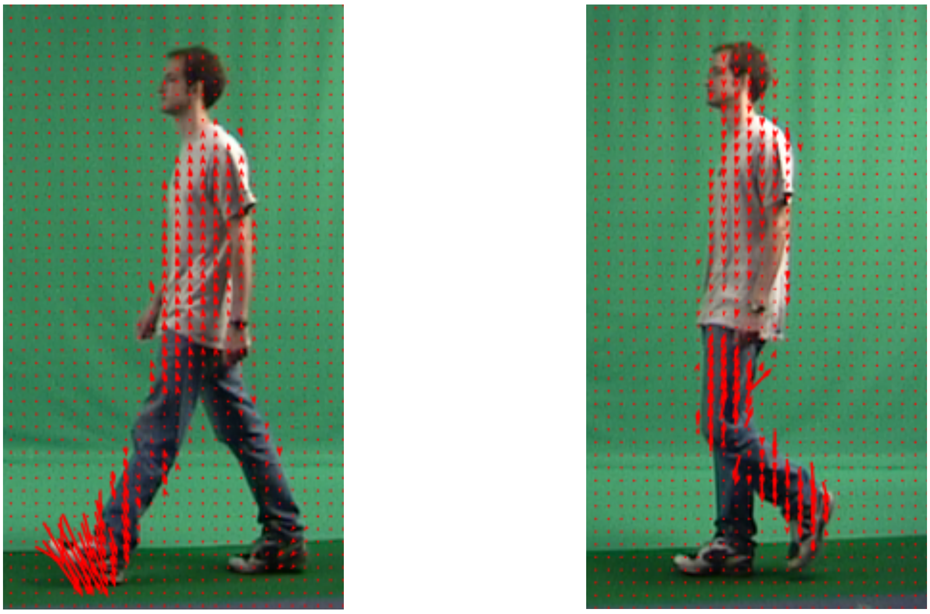


Figure 3.15 The zoom-in radial flow on the walking subject.

The image sequence Dumptruck gives another encouraging example in Figure 3.16. The silver car in the front and the red dump truck in the back are approximately moving in constant velocity for waiting the traffic light. In contrast, the detection result shows the other two cars have passed the intersection and both are accelerating. This example demonstrates that our algorithm can differentiate the accelerating objects from those which are simply moving.

(a) Optical flow

(b) Acceleration

Figure 3.16 The flow fields of Dumptruck.

3.6 Conclusions

Previous research in motion analysis has mostly focused on the displacement between frames, without considering the diversity of motion. There has been two previous study attempted to reveal the acceleration from optical flow however the principle in Dong's algorithm [52] produces isotropic results since the flow fields are too smooth for detecting flow on them. Chen [51] actually measure the jerk field in their work.

First acceleration is derived from the basis of Horn-Schunck, however most real motion violates the basic global smoothness assumptions made in the original formulation. We show another way to approximate acceleration fields which are more accurate and able to handle most situations and appear improved over the Horn-Schunck technique on the standard Yosemite test sequence. The ability of the new algorithms is also demonstrated by its capability to achieve radial and tangential acceleration analysis, providing a completely new way to understand and disambiguate motions in image sequences. Clearly, acceleration is likely to be more sensitive to noise though the experiments show that this is not a severe limitation and in fact radial acceleration error estimates are encouragingly low. In the next chapter, we will demonstrate the application of radial acceleration on gait sequences in order to detect heel strikes.

Chapter 4 Detecting Heel Strikes for Gait Analysis through Acceleration Flow

4.1 Introduction

In gait analysis, heel strikes are an important and preliminary cue for gait analysis because gait period, step and stride length can be derived accurately by the moment and position of heel strikes. It refers to the heel first contacts the ground during the stance phase of the walking cycle [8]. In this chapter, we introduce a new method of using acceleration to detect when and where a heel strikes on the floor. When the foot is approaching to strike, its motion status changes from moving forward to making circular motion centred at the heel. The amount of acceleration on the leading foot will dramatically increase when the heel hits the floor. According to this clue, the key frames can be determined by the quantity of acceleration flow within a Region of Interest (ROI), and positions can be found from the centres of rotation caused by radial acceleration. Compared with previous heel strike detection methods, the temporal template of this new method only requires three consecutive frames for processing so it also allows near real-time detection with only a single frame of delay which able to lead further applications.

Our approach is tested on a number of databases which were recorded indoors and outdoors with multiple views and walking directions for evaluating the detection rate under various environments. Experiments show the ability of our approach for both temporal detection and spatial positioning. The immunity of this new approach to three anticipated types of noises in real CCTV footage is also evaluated in our experiments. The acceleration flow detector is less sensitive to Gaussian white noise, whilst being effective with images of low-resolution and with incomplete body information when compared with other techniques.

This chapter is organised as follows: Section 4.2 describes how we detect heel strike by acceleration and the intuition behind this idea. Section 4.3 introduces the benchmark gait databases used in the experiments. The experiments and analysis of the robustness of our method are presented in Section 4.4. Section 4.5 discusses the current heel strike detection techniques and the advantages and limitation of our algorithm, followed by the summaries in Section 4.6.

4.2 Detecting Heel Strike through Radial Acceleration

4.2.1 The Acceleration Pattern on Gait

Torsos move like connected pendula during walking and researchers have successfully simulated pathological gait by using a liner inverted pendulum model [55], [56]. Pendula have a regular acceleration pattern, which implies that we can describe gait by the acceleration pattern of the image-based data. Figure 4.1 shows the acceleration fields of the body during toe off (a), heel strike (b) and heel rise (c). They reveal that the legs and feet appear to have more acceleration or deceleration than the other parts of the body during different gait phases. Likewise, the forearms have acceleration since they are similar to swinging pendula. Therefore, the acceleration pattern of a walking body could be used to indicate the gait phases. In this work, we detect the heel strike by analysing the radial acceleration of the leading foot [57], [58].

(a) Toe off (b) Heel strike (c) Heel rise

Figure 4.1 The radial acceleration flow on a walking person.

At the instant of a heel strike, the foot hits the ground which forces its velocity to cease in a short time. Therefore, the acceleration of the front foot increases dramatically, due to the disappearance of velocity (rapid deceleration). Also, the striking foot sole's motion is approximately circular during the period between the heels striking on the ground to fully touching the ground, centred at the heel. Hence, most acceleration caused by heel strikes is radial in nature. The video frames where the heel strikes then can be located by the quantity of radial acceleration. When people are walking, the motion of the body is similar to several joined pendula [8]. Therefore, the radial acceleration caused by a heel strike might be confused with that caused by other limbs since the motion of a pendulum incorporates radial acceleration. To reduce interference, we extract the Region of Interest (ROI) which is located on the leading foot according to a walking body model. The size of the ROI is 0.133H × 0.177H where H represents the height, shown in Figure 4.2.

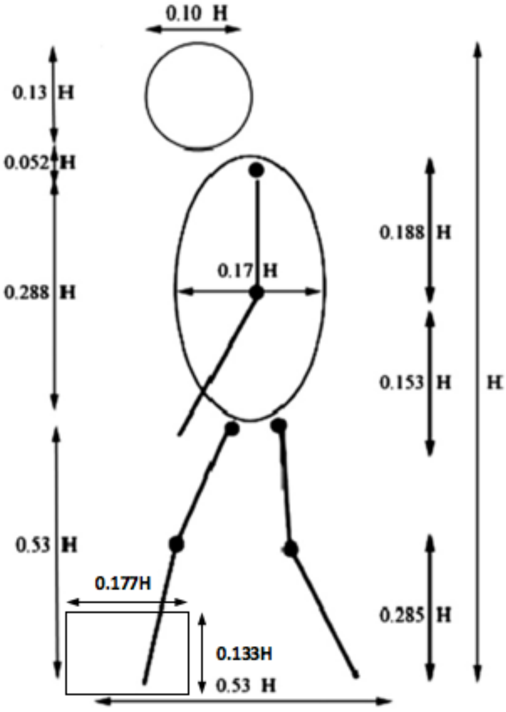


Figure 4.2 Gait Proportions [59].

4.2.2 Strike Position Estimation and Verification

After obtaining the acceleration field $\mathrm{A}(t)$, we can use the algorithm F which decomposes resultant acceleration into radial and tangential acceleration to compute the tangential acceleration field $\mathrm{T}(t)$, radial acceleration fields $\mathrm{R}(t)$ and the radial acceleration rotation centres map $\mathrm{C}(t)$:

$$\mathrm{F}(\mathrm{A}(t)) = \begin{cases} \mathrm{T}(t) \\ \mathrm{R}(t) \\ \mathrm{C}(t) \end{cases} \tag{4.1}$$

where $\mathrm{A}(t)$ is the acceleration field at time t.

If all the detected radial accelerations in the ROI are caused by the circular motion of the foot, then the rotation centres of these radial accelerations should all locate at the heel. Consequently, the densest point of the rotation centres map indicates the strike position. There are numerous algorithms to determine the densest point and we selected three methods to experiment in this thesis: weighted sum, accumulation proximity [5] and mean shift [60].

A weighted sum is a straightforward method to estimate where the rotation centres accumulate. The strike position is determined by:

$$h(t) = \frac{\sum_{i,j}^{m,n} w_{i,j} \times (i,j)}{\sum_{i,j}^{m,n} w_{i,j}} \tag{4.2}$$

where the weight factor is determined by the density of point (i,j) in centres map $\mathrm{C}(t)$:

$$w_{i,j} = \mathrm{C}(t)_{i,j} \tag{4.3}$$

Bouchrika and Nixon [5] accumulated all the corners in the gait sequences into one image. A dense area is the strike location since the striking foot is stabilized for half a gait cycle. For estimating the dense area, they estimated a measure for density of proximity. The value of proximity at point (i,j) is determined by the number of corners within the surrounding region. By using the same method, the density of proximity for rotation centres map $\mathrm{C}(t)$ is estimated for evaluation in this thesis. If $\mathrm{R}_{i,j}$ is a sub-region in the ROI with centre (i,j), the proximity value $d_{i,j}$ is computed by:

$$\begin{cases} d_{i,j}^{r} = \dfrac{\mathrm{C}_r}{r} \\ d_{i,j}^{n-1} = d_{i,j}^{n} + \dfrac{\mathrm{C}_n}{n} \end{cases} \tag{4.4}$$

where r is the radius of sub-region $\mathrm{R}_{i,j}$ which is around 20 pixels, $d_{i,j}^{n}$ is the proximity value for rings which are of n pixels away from the centre (i,j), $\mathrm{c_n}$ is the sum of rotation centres in the centre density map $\mathrm{C}(t)$ for rings which are of n pixels away from the current processing centre (i,j). Equation (4.4) is repeated for each point in the ROI to obtain the rotation centres density of proximity. The densest point in the density of proximity is where the heel strikes the ground.

Mean shift [60] is a recursive algorithm that allows nonparametric mode-based clustering. It assigns the data points to the cluster iteratively by shifting points towards the highest density of data points. The algorithm of iteratively finding the highest density in mean shift is used for positioning the densest point in the rotation centre map. Since we only need one mode in the ROI, the bandwidth needs to be wide. For the experiments in this thesis a value of 20 pixels is used. The comparison experiments will be illustrated in Section 4.4. Before analysing the experiments, we are going to introduce the databases in the next section.

4.3 Gait Databases

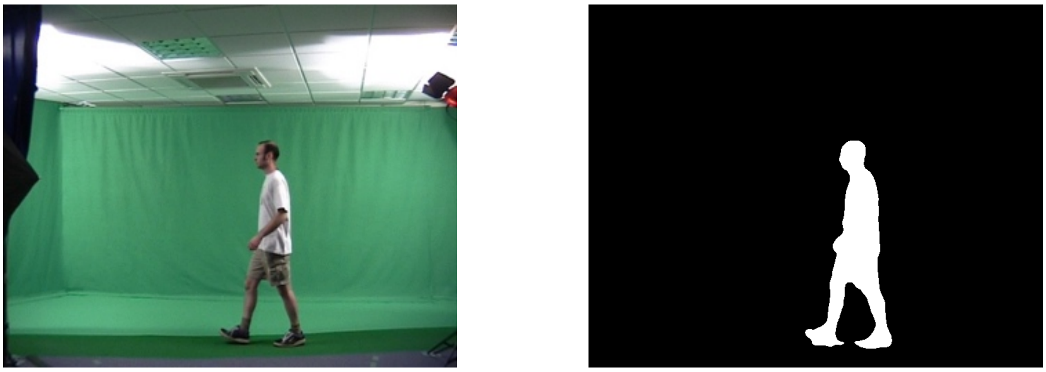
In the above sections, we propose a new technique to detection heel strike for gait analysis. We evaluate our method on three gait benchmark databases: the Large Gait Database (SOTON) [54], the CASIA Gait Database (CASIA) [61]–[63] and the OU-ISIR Gait Database (OU-ISIR) [64], [65]. The covariates and number of subjects of each database are illustrated in Table 4.1. Here we introduce them.

Database	# of subjects	Data covariates
SOTON	115	2 views, 3 scenarios (indoor/outdoor track, treadmill)
CASIA-A	20	3 views
CASIA-B	124	11 views, clothing, baggage
OU-ISIR	10,307	14 views

Table 4.1 The summary of gait data used in the experiments.

4.3.1 The Large Gait Database

The Large Gait Database (SOTON) [54] was built in 2002 by the University of Southampton. Shutler et al. collected walking sequences from over 100 subjects, including indoor (controlled lighting) and outdoor (uncontrolled lighting). Each target has 8 sequences of approximately 1.5 gait periods. This database focuses on colour and lighting quality, they used two different types of cameras and captured 3 scenarios each from 2 different views. The video frame rate is 25fps. Since the database focussed on the key factors that affect gait recognition other than background segmentation, Chroma-keying was used in indoor data recording. Chroma-keying is segmenting a relatively narrow range of colour, green was used because it is further away from skin tone and is less likely to be a component of clothing. This technique helps to remove the background from the subject in the images. Figure 4.3 shows some examples from the SOTON database and silhouettes.



(a) Inside data (walking to the left)

(b) Silhouette of (a)

(c) Inside data (walking to the right)

(d) Outdoor data

Figure 4.3 Examples of data in SOTON Large Database, it contains indoor and outdoor data with two walking directions, and their processed silhouettes.

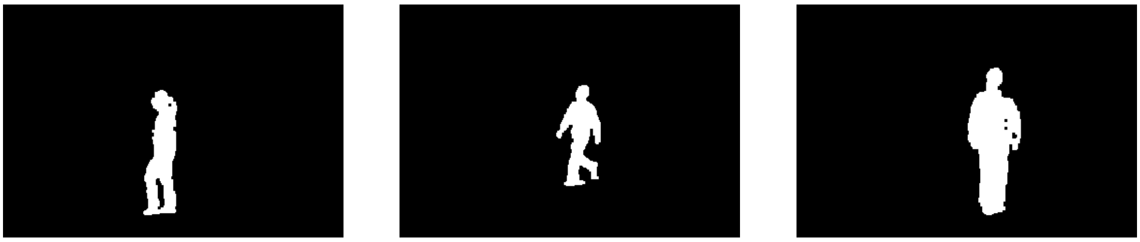
4.3.2 CASIA Gait Database

The CASIA Gait Database Dataset A (CASIA-A) [61] and Dataset B (CASIA-B) [62], [63] were built by Institute of Automation, Chinese Academy of Sciences in 2003 and 2006 separately. Before CASIA, there were only a few databases designed for gait recognition and most of them only contain a small number of subjects and walking environments which limited the progress of the field. The emergence of this database revealed the key factors of gait recognition and offered a baseline for new algorithms.

CASIA-A contains 20 subjects walking in three directions (0°, 45°, 90°) with respect to the camera in an outdoor environment. Four sequences were filmed for each view per subject. These image sequences are in 24-bit with resolution 351 × 240 at frame rate 25 fps. Each sequence contains 90 frames on average.

The videos in CASIA-B were all recorded in an indoor environment. This dataset contains 124 subjects, 93 males and 31 females, in three different categories of data: normal walk, wearing coats and carrying bags. The subjects were walking along a specified trajectory enclosed with 11 cameras at different angles. Figure 4.8 illustrates some sample frames and processed silhouettes in CAISA-A and CASIA-B with multiple walking directions.

(a) CASIA-A (0°) (b) CASIA-A (45°) (c) CASIA-A (90°)



(d) Silhouette of (a) (e) Silhouette of (b) (f) Silhouette of (c)

(g) CASIA-B (18°) (h) CASIA-B (144°) (i) CASIA-B (162°)

(j) Silhouette of (g) (k) Silhouette of (h) (l) Silhouette of (i)

Figure 4.4 Data of different views and processed silhouettes from CASIA-A and CAISA-B.

Figure 4.4 shows some examples from CASIA-A and CASIA-B. The images in the first row are gait sequences from CASIA-A at angle 0°, 45° and 90° and second row are their extracted silhouettes. The third row are data from CASIA-B at angle 18°, 144° and 162° respectively and the fourth row are their silhouettes.

4.3.3 The OU-ISIR Gait Database

The OU-ISIR Gait Database (OU-ISIR) [64], [65] 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 [65], which is two times larger than their earlier version in 2012 [64]. The huge amount of data significantly benefits the applications of machine learning algorithm gait recognition that emerged in recent years. Moreover, the diversity of the subjects' age and gender leads to statistically reliable performance evaluation of gait algorithms.

Figure 4.5 Examples of subjects in OU-ISIR, the gender and age are widely distributed [64].

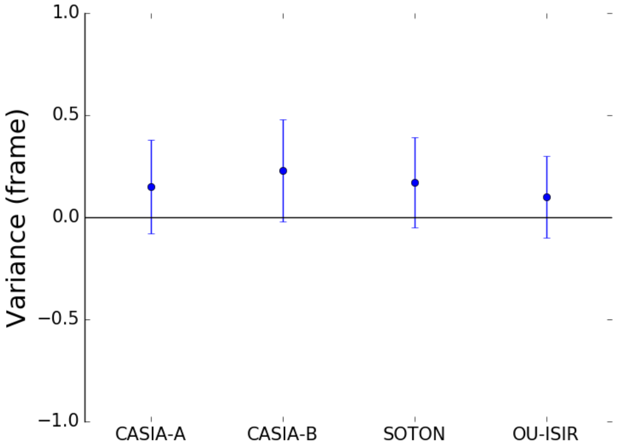
4.4 Experimental Results

We evaluate our heel strike detection method on three benchmark databases: CASIA [66], [67], SOTON [54] and OU-ISIR [64]. The data used in this thesis is collected with various controlled environments. We test around 100 heel strikes in each scenario and the test data incorporates multiple viewing angles and walking directions with gait sequences recorded indoors and outdoors, as described in Table 4.2. The acceleration decomposition algorithm is based on a subject moving perpendicular to the background so it is theoretically most effective in a direction perpendicular to the camera. Therefore, gait data imaged at multiple views has been used to evaluate the robustness of our approaches to other view angles.

Databases	CASIA-A (45°)	CASIA-A (90°)	CASIA-B	SOTON	OU-ISIR
Lighting Control	NO	NO	Yes	Yes	Yes
Camera Visual Angle (°)	45	90	54	90	~75
Number of subjects	13	25	15	21	15
Number of strikes	96	98	126	114	120
Frame size	240×352	240×352	240×320	576×720	480×640

Table 4.2 Database information.

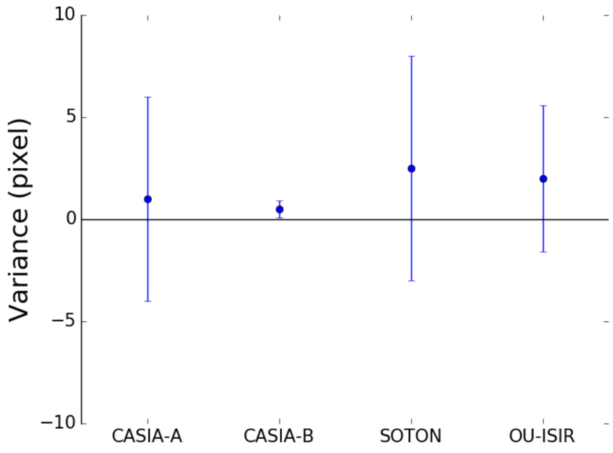
The GT of key frames and heel strike positions were manually labelled multiple times by three different people. Figure 4.6 shows the variance of manually labelled GT between different databases for key frames and strike positions. The variance in the key frame labelling is generally low and within one frame. Figure 4.6 (b) shows greater variance on the SOTON dataset as it has the largest the ROI compared with other databases. For further research and evaluation, the implementation code and heel strikes' GT are publicly available[7].



(a) Key frame

---

[7] https://github.com/YanSunSoton/HeelStrikeAcc.

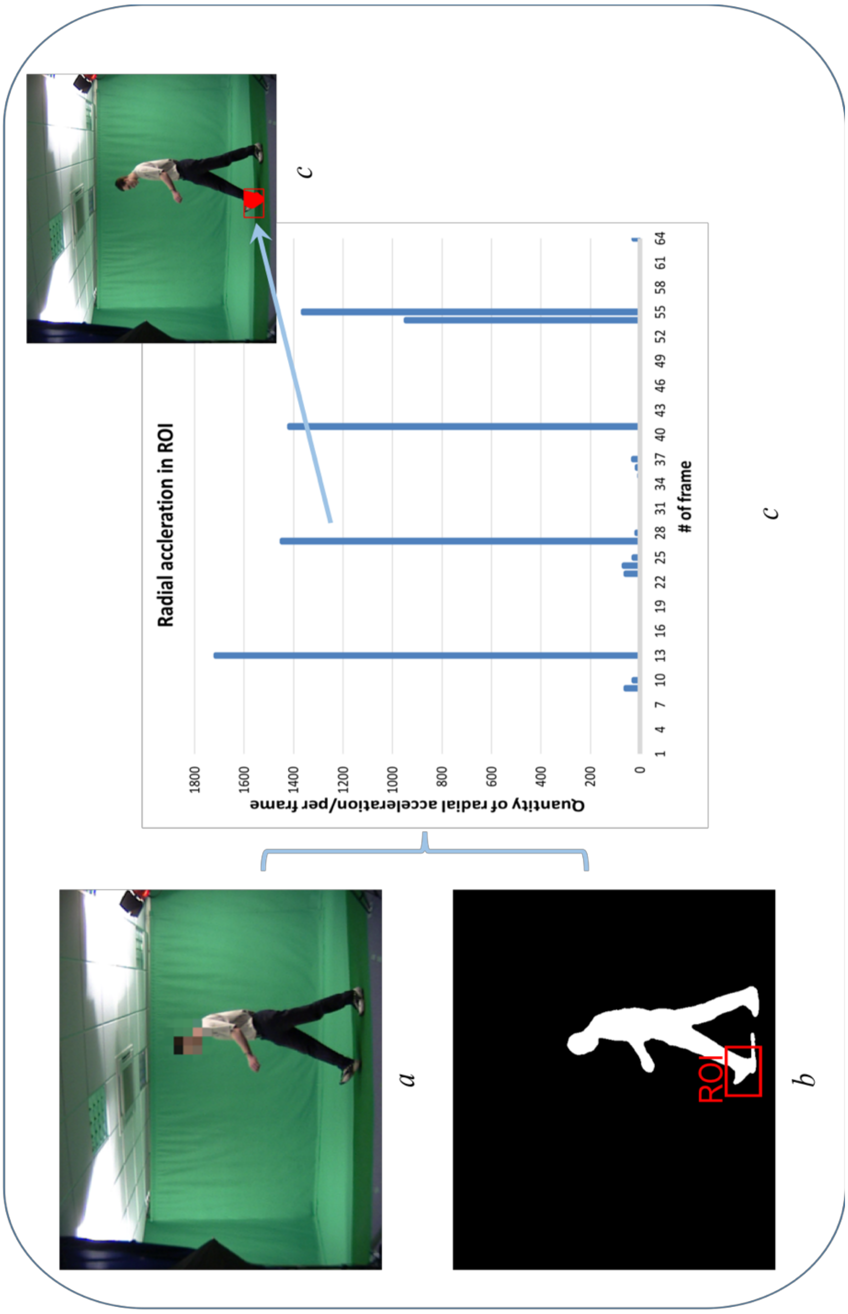


(b) Heel strike position

Figure 4.6 GT labelling variance on different databases.

4.4.1 Key Frame Detection

The key frame (or moment) of a heel strike is detected according to the quantity of radial acceleration in the ROI as described in Section 4.2.1. The histogram of radial acceleration within a walking sequence shows distinct suggestions for key frames, it appears regularly and noticeably, showing the periodicity of gait.



(a) The gait sequences.

(b) Extracted silhouettes and derived the ROI.

(c) Detected radial acceleration in the ROI.

Figure 4.7 An overview of key frame detection.

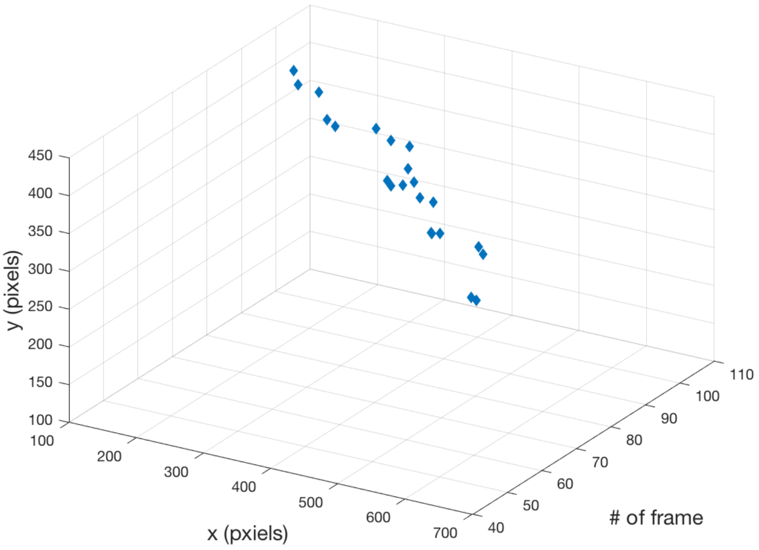
The framework of our new heel strike detection system is illustrated in Figure 4.7: (a) illustrates an example of heel strike key frame and (b) is the area that we compute dense radial acceleration flow, (c) shows the histograms of radial acceleration within the ROI during a gait video. Noise has been effectively reduced by applying an empirical threshold. In (c), the heel strikes occur at frame 13, 27, 41 obviously. There is much acceleration flow in frames 54 and 55 since the heel strike took place between them. This suggests that a higher frame rate could improve accuracy of detection. Figure 4.8 gives the pseudo code of the system.

```
for frame in video:
      vel_field_1 = DeepFlow(frame_2, frame_1)
      vel_field_2 = DeepFlow(frame_2, frame_3)

      acc_field = vel_field_1 + vel_field_2
      rad_field, centre_map = decomp_components(acc_field)

      ROI = extract_region(silhouette_2)
      rotation_center = desity_accumulation(centre_map[ROI])

      for each_pixel in ROI:
            if rad_field[each_pixel] > magnitude_thres:
                  rad_amount += 1
            else:
                  pass
      end for

      while rad_amount is peak:
            "KEY FRAME!"
            strike_position = rotation_center
      end while
end for

return key_frames_num, strike_positions
```

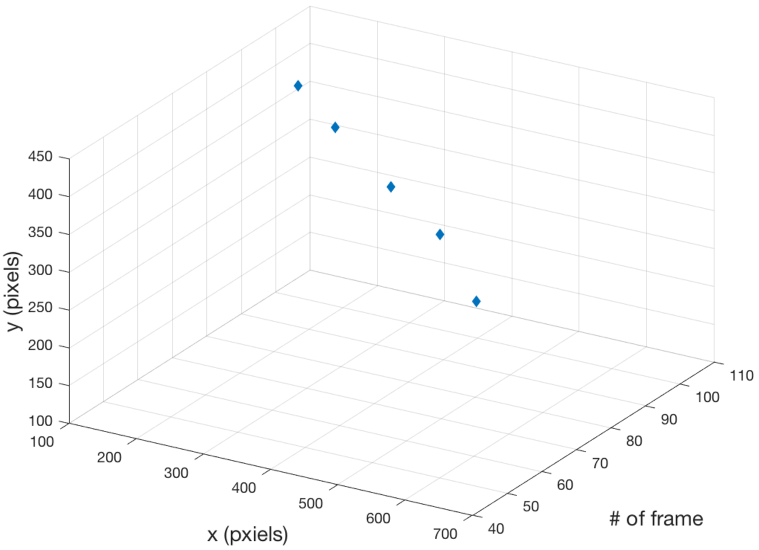
Figure 4.8 Heel Strike Detection System.

4.4.2 Heel Strike Position Verification

The ROI is extracted according to gait proportions is not always perfectly located on the leading foot in the sequence because the shape of the human body changes during a gait cycle. In addition, there is radial acceleration on the other body parts, for example the calf, since the limbs' motion is that of several joined pendulums [8]. The rotation centres of these erroneous radial accelerations also form invalid strike position candidates. To reduce the effect of this error, the detected key frames are used to filter the heel strike position candidates. When the heel strikes between two frames, the acceleration quantities are used as a weighting factor for deriving the positions. Figure 4.9 (a) shows detected candidates of heel strike positions in each frame and (b) is the result after being filtered by key frames. The (expected) periodicity of gait is evident in the result.



(a) Candidates for heel strikes.



(b) Detected heel strikes (after filtering).

Figure 4.9 Heel strike verification process.

4.4.3 Detection Performance

Bouchrika and Nixon proposed a method that accumulated corners within a gait cycle by the Harris corner detector to determine the positions of heel strikes. Theoretically, there should be dense corners accumulated at the positions of heel since the heels stay at the strike positions for almost half gait cycle [5]. We compare our detection results against the corner detection method since there are few heel strike detection methods based on standard image sequences with implementation available. The performance is evaluated by F-score:

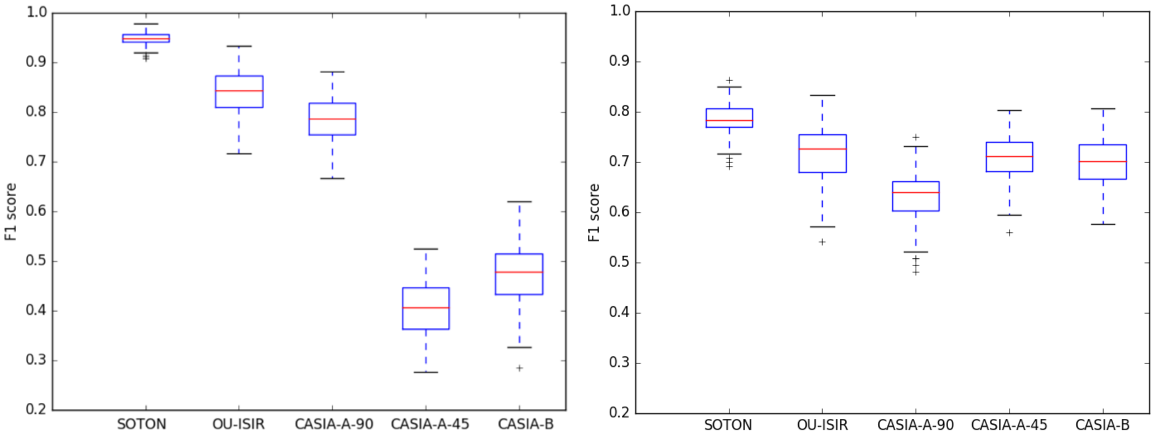
$$\mathrm{F}_{\beta} = (1+\beta^2)\frac{p * r}{r+\beta^2 p} \tag{4.5}$$

p stands for precision and r for recall. F score prefers precision if β is set to be small and recall if β is large [68]. Let TP represent the number of true positives, TN the true negatives, FP is false positives and FN is false negatives, p and r can be computed by:

$$p = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \quad r = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{4.6}$$

Figure 4.10 and Figure 4.11 illustrate the comparative F1 score of acceleration and corner detector. Since F-score favours precision if β is small and recall if β is large, β is set as 1 here to evenly balance the result. The results differ from the earlier results [50] because the background has been included to give a more realistic implementation scenario. The detection of the heel strike moments (the key frames) and positions are evaluated separately since they are determined individually, and they describe different events in gait analysis.

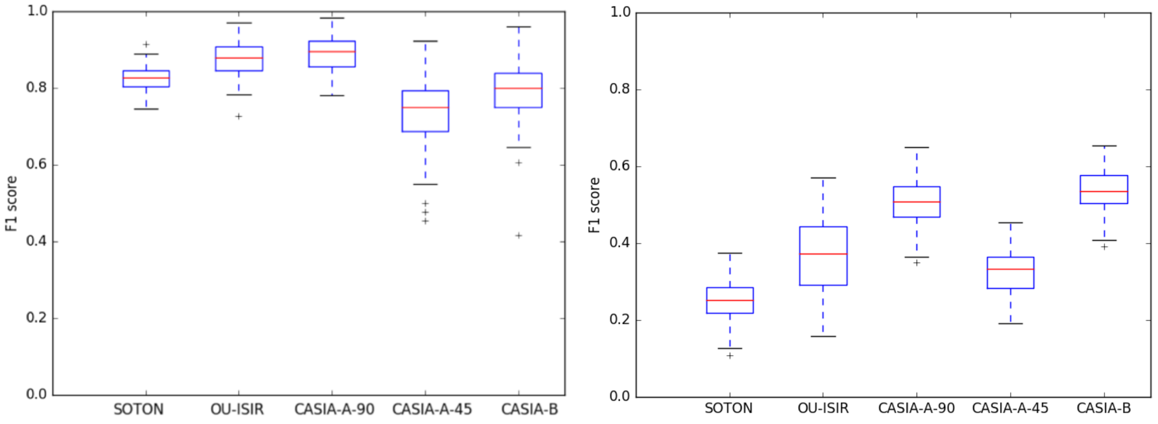
Since the corner detection does not return the key frames, an additional condition is applied which is that for a key frame to be successfully detected, a corner position within ±30 pixels from the GT is considered as a true positive. This condition is actually quite generous and leads to an optimistic estimate of the frame for corner detection. For the radial acceleration detector, the criterion for a true positive in Figure 4.10 is if the detected frames are within ±2 frames from the GT. For heel positioning in of both methods, a distance within ±10 pixels (along both axes) from the GT is considered as a true positive in Figure 4.11.



(a) Radial acceleration

(b) Harris corner detector [5]

Figure 4.10 F1 score of key frame detection.



(a) Radial acceleration

(b) Harris corner detector [5]

Figure 4.11 F1 score of heel positioning.

Figure 4.10 illustrates that radial acceleration is able to detect key frames accurately when the camera is nearly perpendicular to the walking direction, the detection rate decreases with the increase of the angle between the camera and the walking subjects. Acceleration is more sensitive to the view angle since the scale of acceleration changes through the image sequence if the walking trajectory is not perpendicular to the camera. When the angle is large, the magnitude of acceleration is extremely small if it is far from the camera, which will cause the failures. Figure 4.13 shows that the lost most occur when subject is far from the camera.

In Figure 4.11 the radial acceleration detector provides more precise positioning results than Harris corner detector for all the camera views, especially on SOTON. The main reason is that the image size of SOTON is large, which causes the excessive accumulation area for Harris corner detector, thus the precision decreases correspondingly.

The Precision-Recall curves respect to the magnitude of acceleration (varies from 0 to 4) and the density of corners (varies from 500 to 1600) are reported in Figure 4.12. Since the strike positions are filtered by the key frames they did not show significant sensitivity to the change of threshold, hence only the algorithm of key frames detection is evaluated.

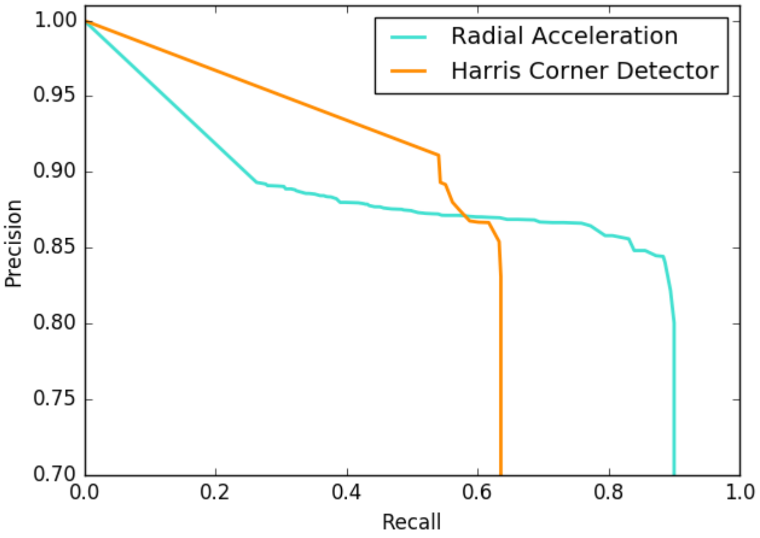


Figure 4.12 The key frame detection Precision-Recall curves of radial acceleration and Harris corner detector.

Radial acceleration hits a much higher recall rate therefore it has a larger area under the curve than corner detector. The precision of radial acceleration is steady, it keeps around 86% with the change of recall while the PR curve shows that the corner detector is more sensitive to the density.

Figure 4.13 shows samples of the detection results for different databases. Our algorithm can locate a precise location and frame of the heel strike since the angle between camera and walking subject is small. In CASIA-A-45 (the last row) the acceleration detector failed to detect several strikes when the subjects walk away from the camera and the accuracy of localization also decreases.

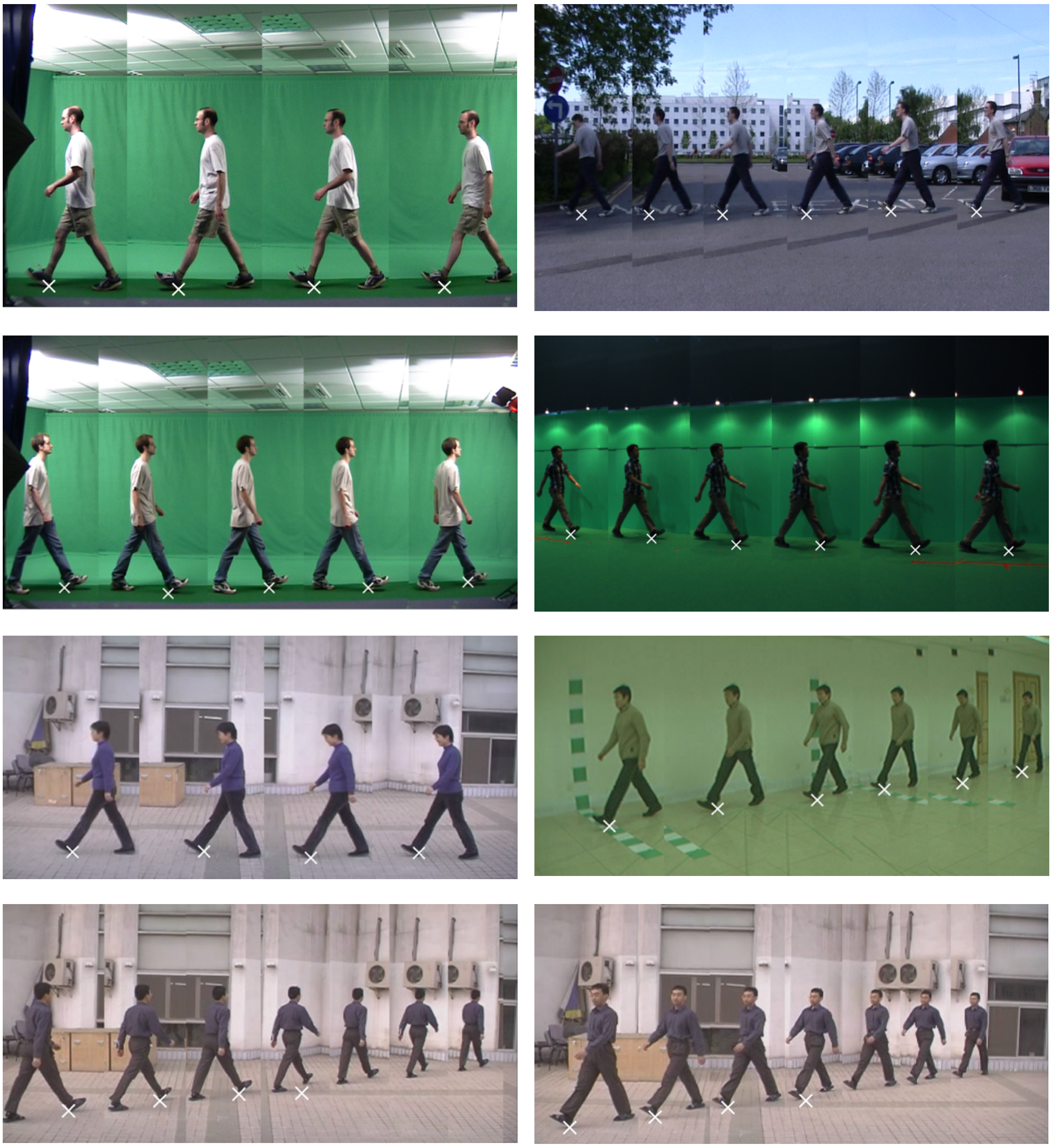


Figure 4.13 Examples of detection results with various databases.

Positioning is critical for the heel strike detection in gait analysis. In the above experiments, the weighted sum is used to estimate the strike positions. For improving the performance, a measure for density of proximity by Bouchrika [28], and, mean shift are also tested. The algorithm and parameters have been explained in Section 4.2.2: we use weighted sum, density of proximity and Mean shift to determine the strike position. There is only one strike position in the ROI so one mode only needs to be determined, the bandwidth is set as 20 pixels in the experiment. Around 100 randomly selected strike frames from SOTON database are tested. The new condition of a correct prediction in Table 4.3 is the estimated position is within ±3 pixels from the GT. Mean shift has improved the precision of positioning significantly considering the condition of initial results in

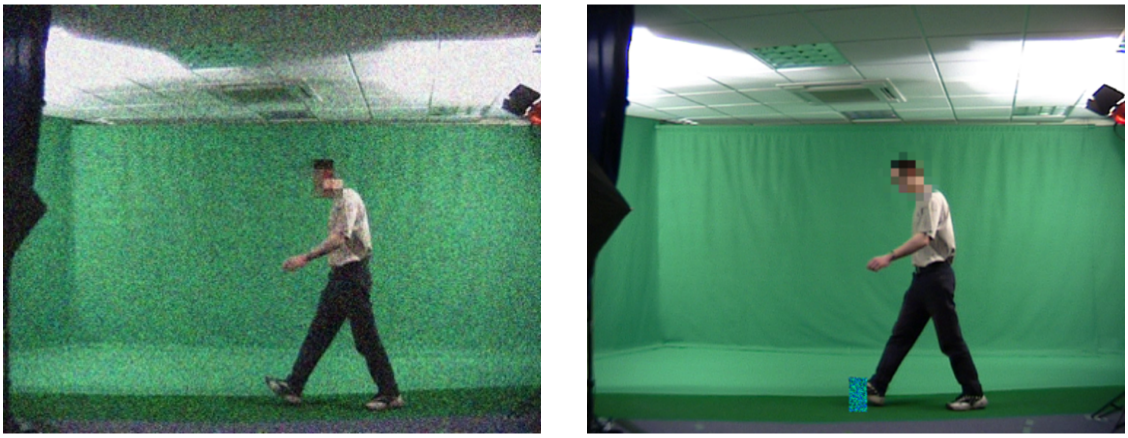
Figure 4.11 (a) was within ±10 pixels. The great improvement of determining the strike position by density of proximity [28] via acceleration also proves that the radial acceleration is a much better feature for detecting heel strike than corners.

Method	F1-score
Density of proximity [5]	0.72
Mean shift	0.95

Table 4.3 The results of strike position detection by advanced algorithms.

4.4.4 Robustness of Heel Strike Detection Approaches

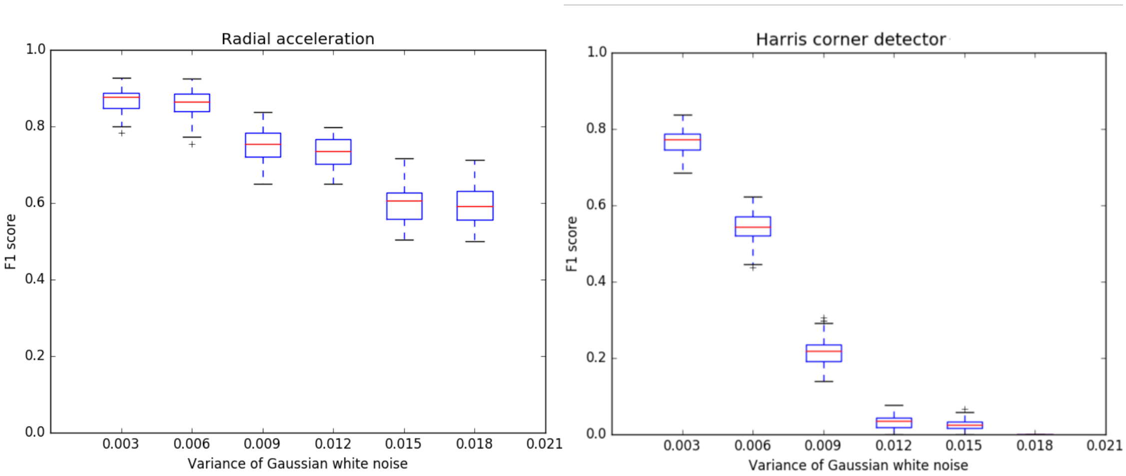
Since the performance of a system under adverse imaging conditions is an important issue, we evaluate the robustness of our heel strike detection technique. Three different factors affecting image quality, that might reduce the detection rates, are applied to the original sequences: Gaussian zero-mean white noise, occlusion in the detection area, and reduced resolution. These factors reflect some of the difficulties anticipated when detecting heel strikes in real surveillance videos. Figure 4.14 illustrates the examples of different types of noise at different levels.



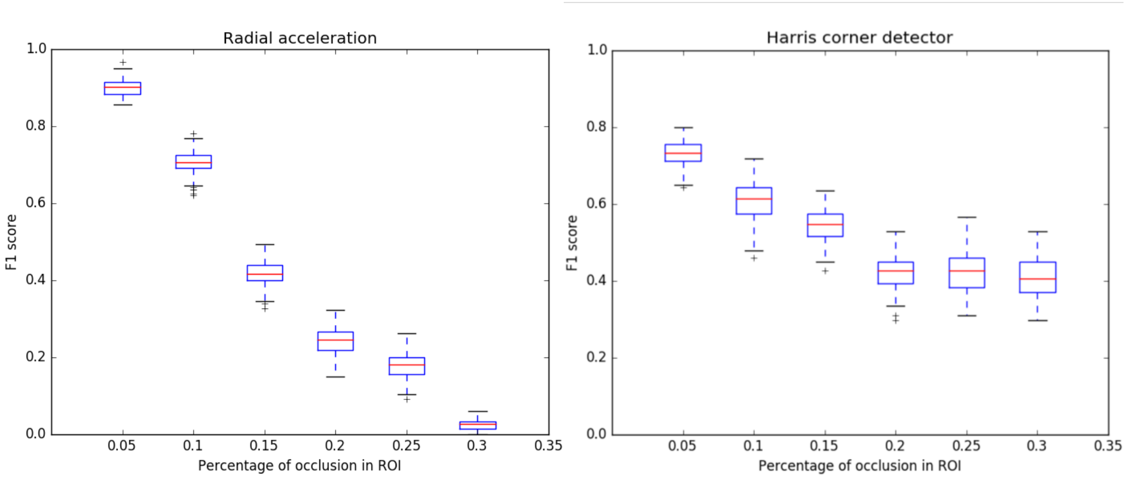
(a) Gaussian white noise (=1.5%)

(b) Occlusion in the ROI (40%)

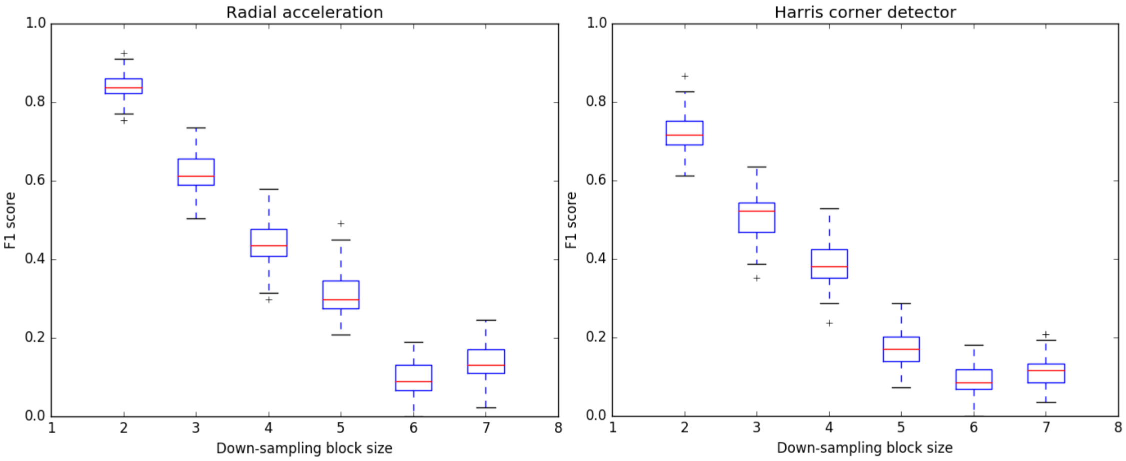
Figure 4.14 Example of added noise and occlusion.



(a) Testing immunity to Gaussian white noise.



(b) Testing immunity to occlusion.



(c) Testing immunity to resolution.

Figure 4.15 Performance analysis of heel strike detection.

Figure 4.15 gives the results of testing the acceleration detector’s immunity to these factors. Corner detection is also evaluated for comparison. The performance of the acceleration detector reduces slowly and smoothly with increased Gaussian noise variance. Corner detection is much more sensitive to increases in variance of the Gaussian noise, as shown in (a).

The evaluation of immunity to occlusion investigates whether the gait information in real surveillance can be totally seen, or not. Occlusion was achieved by covering the ROI from the toe to the heel increasingly with a random texture. The performance under occlusion decreases steadily, and our approach failed when the ROI is covered more than 30% of the whole area. This is because most high-magnitude acceleration is located around the toe (the toe travels the greatest distance during heel strikes) but the toe is almost completely occluded when the occlusion in the ROI is over 30%. The detection by corners does not decrease significantly since the area that most corners are concentrated on the heel which has not been totally occluded yet. Acceleration does outperform the Corner method when occlusion is slight. If the occlusion were to start from heel to toe, our method could achieve much better immunity than the corner detector.

Resolution reduction investigates whether resolution of the subject is sufficient in surveillance footage. The original images are down-sampled and the detection rate of both approaches decreases to a low level when the new pixels are equivalent to 5×5 patches in the original image (in which the height of the subject is now around 70 pixels whereas it was 350 pixels originally). Acceleration and corner detectors show similar immunity characteristics under this situation.

4.5 Discussion

In dynamics, a change of force causes acceleration, and acceleration changes motion. Consequently, acceleration is a distinctive cue to the change of motion. Some previous physics-based gait analysis approaches used accelerometers and gyroscopes to detect the acceleration and angular velocity of the body parts in order to determine stance, swing and strike [1], [3]. We have applied this principle to standard image sequences to detect heel strikes. When the heel approaches the strike, the foot has significant radial acceleration which is centred at the heel. Only the acceleration flow in the region around the leading foot is considered in our approach to reduce the effects of noise. Our experimental results show that acceleration is a more powerful way of estimating the positions of the strike than previous standard image-based techniques. Also, our method overcomes the problem of detection in real-time as only three frames are needed for estimating acceleration flow. The evaluation of immunity to different types of noise suggests that acceleration is more robust to Gaussian noise than the previous approach.

On the other hand, the main limitation of acceleration is its sensitivity to the visual angle between the camera and the subject. When the camera is orthogonal to the subject, acceleration performs best since the measurements and decomposition algorithms utilise a 2-D plane. The most realistic way of solving this problem is by applying the algorithm using a 3-D volume, for example using Kinect depth images, to replace the standard image sequence. However, the complexity of computation will be much higher than the existing technique.

Another weakness of this approach is that it can only be applied in data with a static background and the subjects in the images do not overlap. This is a similar limitation to most existing techniques however. Currently background subtraction and silhouette extraction are essential pre-processing progresses for most standard image-based approaches. The results will be severely affected if the scene is too complex, for example the overlap of subjects in a crowded scene. Hence there is still refinement necessary to be able to apply these techniques in poor quality images, such as surveillance footage of the underground, or to videos recorded with adverse illumination.

4.6 Conclusions

This chapter demonstrates that acceleration flow can be used for detecting heel strikes. Cunado and et al. [8] proposes that the limbs appear to have pendulum-like motion in their gait model and acceleration has been widely used in gait analyse techniques based on physics data as the motion of a pendulum can be easily discriminated by radial and tangential acceleration. The ability of the new heel strike detection technique has been compared with one of the few existing techniques. The results show that this new technique not only improves the precision significantly but also enables real-time detection. By using mean shift to localize the strike positions base on the rotation centres map, the precision reaches 95%. The experiments also investigate how camera viewpoint can affect the performance, as radial and tangential components are derived based on a plane perpendicular to the subjects. After using radial acceleration to detect heel strike, we are going to explore the higher order of motion in the next chapter.

Chapter 5 Jerk and Higher Order Motion in Computer Image Streams

5.1 Introduction

After successfully separating acceleration from complex motion and applying it into heel strike detection, we now investigate the higher orders of motion. Experimental results on synthetic images show that higher order flow gives a different perspective from acceleration, and they allow more possibilities for analysing complex motion fields in computer vision. Following the detection system in the last chapter, jerk and snap are also applied for detecting heel strikes.

This chapter is arranged as follows: Section 5.2 introduces the physical definitions of jerk and snap separately, including their applications in industry. The algorithms are presented in section 5.3. Section 5.4 illustrates various motion fields both on the synthetic and real-world image sequences; in addition the higher orders of motion flow are applied in the heel strike detection system. Discussions and conclusions can be found in Section 5.5.

5.2 Jerk, Snap and Higher Order Motion in Kinematics

It is well known that velocity measures the change in position over time, and acceleration is the change of velocity, the term of measuring the change of acceleration is Jerk [69], [70]. It is usually used for analysing chaotic dynamical systems [71]. Acceleration links a force acting on a mass from Newton's Second Law:

$$\vec{F} = m\vec{a} \tag{5.1}$$

Hence, assuming constant mass, jerk describes the change of force; and Snap describe the change of jerk in kinematic. In calculus snap is the derivative of acceleration with respect to time, the fourth derivative to position [70]. Equation (5.2) describes the time evolution of position $\vec{r}$:

$$\vec{s}(t) = \frac{d\vec{j}(t)}{dt} = \frac{d^2\vec{a}(t)}{dt^2} = \frac{d^3\vec{v}(t)}{dt^3} = \frac{d^4\vec{r}(t)}{dt^4} \tag{5.2}$$

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 nth-order flow under limited snap is shown in Figure 5.1 which gives a sense of the relationship between them.

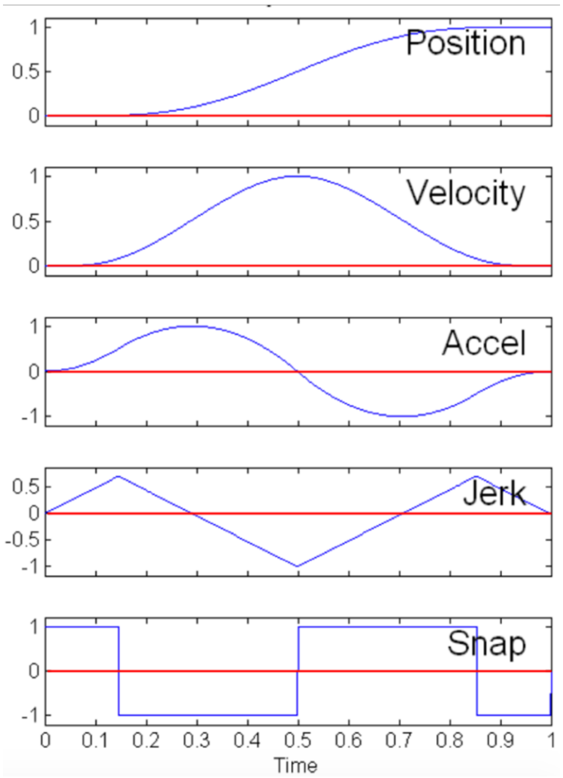


Figure 5.1 The relationship between motion profiles in a straight linear motion [72].

The conventional engineering application of jerk and snap in industry is in motion control since humans have limited tolerance to the change of force. Motion limitation is therefore necessary to avoid users losing control during transportation, $2.0\ m/s^{-3}$ in a straight-line transportation is acceptable for most people. The most common examples are in lift 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 passages' comfort in autonomous driving system [5], [6]. 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 [76].

In road and track design, unbounded radial jerk needs to be avoided on curved parts: the theoretical optimum strategy is linearly increasing the radial acceleration. Another application using acceleration and jerk is the operation path evaluation of numerical control machines [77]. The square of the magnitude of jerk integrated over time termed the "Jerk cost", was measured for quantitatively analysing different movements in human arms [78]. In 2006, Caligiuri and et al. used jerk to monitor how the drug-induced side effect affect patients' handwriting [79]. More recently, a detection algorithm for manoeuvring targets using radar has considered analysing jerk [80], [81].

5.3 Jerk and Snap Fields Estimation

Previous research which has considered jerk mostly derives jerk from the trajectories of moving objects by tracking. Datta and et al. [82] computed acceleration and jerk vectors from the objects' head, which was tracked by the colour sum of squared differences, for person-on-person violence detection. Zaki and et al. [83] get the vehicles' jerk from the trajectory extracted by Kanade-Lucas-Tomasi Feature Tracker algorithm [84]. To the best of our knowledge, we are not aware of any previous computer vision research that has considered analysis of the general acceleration field in images beyond its basic form.

Following our algorithm of Differential-Acceleration, jerk is computed by differencing the neighbour acceleration fields:

$$\hat{\mathbf{J}}(t) = \widehat{\mathbf{A}}\big(t\sim(t+\Delta t)\big) - \widehat{\mathbf{A}}\big((t-\Delta t)\sim t\big) \tag{5.3}$$

where $t\sim(t+\Delta t)$ represent the flow field from frame t to $(t+\Delta t)$. The jerk field is resolved into tangential and radial components as well, they are computed in the same manner with Equation (5.3). The definitions of the tangential and radial components of jerk in this thesis indicate that they measure the linear changes of tangential and radial components of acceleration.

In Differential-Acceleration, we are able to refer the displacement to the middle frame to avoid the inconsistent start positions since it requires three points to compute acceleration. Estimating snap involves an odd number of positions (five), which makes referring the middle point as the start possible. Therefore, snap fields are computed in a similar manner to Differential-Acceleration in Section 3.4:

$$\hat{\mathbf{S}}(t) = \hat{\mathbf{J}}\big(t\sim(t+\Delta t)\big) - \big\{-\hat{\mathbf{J}}\big(t\sim(t-\Delta t)\big)\big\} \tag{5.4}$$

Now we have the algorithms to provide the change of acceleration and jerk, in the next section they are going to be applied on synthetic and real images to see whether they can reveal different motion features.

5.4 Appling Multi-orders Flow Fields on Synthetic and Real Images

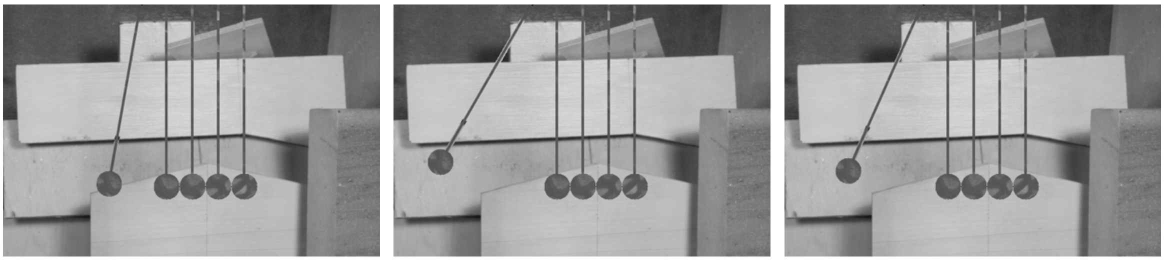
5.4.1 Experiments

The advantages of synthetic images have been mentioned in Section 3.4.2. We manipulate a set of images to simulate the motion of Newton's cradle named "synthetic cradle". It is a device consisting of a set of swinging spheres, which was originally used for demonstrating conservation of momentum and energy. Newton's cradle is a good example for demonstrating the difference between various flow fields due to the different order components change from time to time whereas the entire motion is rather simple.

In the image sequence, the highest point is considered as the stationary point, which is $t = 0$. Swinging to the lowest point is the positive direction and to the stationary position is negative. The change in inclination of the line suspending the swinging ball is computed by:

$$\Delta\theta = 2 * t \tag{5.5}$$

where $t \in \{-5, -4, \dots, 4, 5\}$, $\Delta\theta$ is the degree increased between each frame. Figure 5.2 gives the examples of $t = -5, 0, 2$.



(a) $t = -5$ (b) $t = 0$ (c) $t = 2$

Figure 5.2 Examples of "synthetic cradle".

The results of nth order motion flow when $t = -1, 0, 1$ are presented in Figure 5.3. When $t = -1$, the sphere is swinging to the highest position and remains there for one frame ($t = 0$), then it swings back at $t = 1$. The acceleration shows similar motion fields 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 have the same direction with acceleration and the magnitude of snap is larger, which means that the acceleration is increasing with the same direction. In the second-row jerk and snap flow have notable reduction compare with in frame -1. When $t = 1$, the directions of jerk and snap are opposite which proved that the acceleration is actually decreasing.

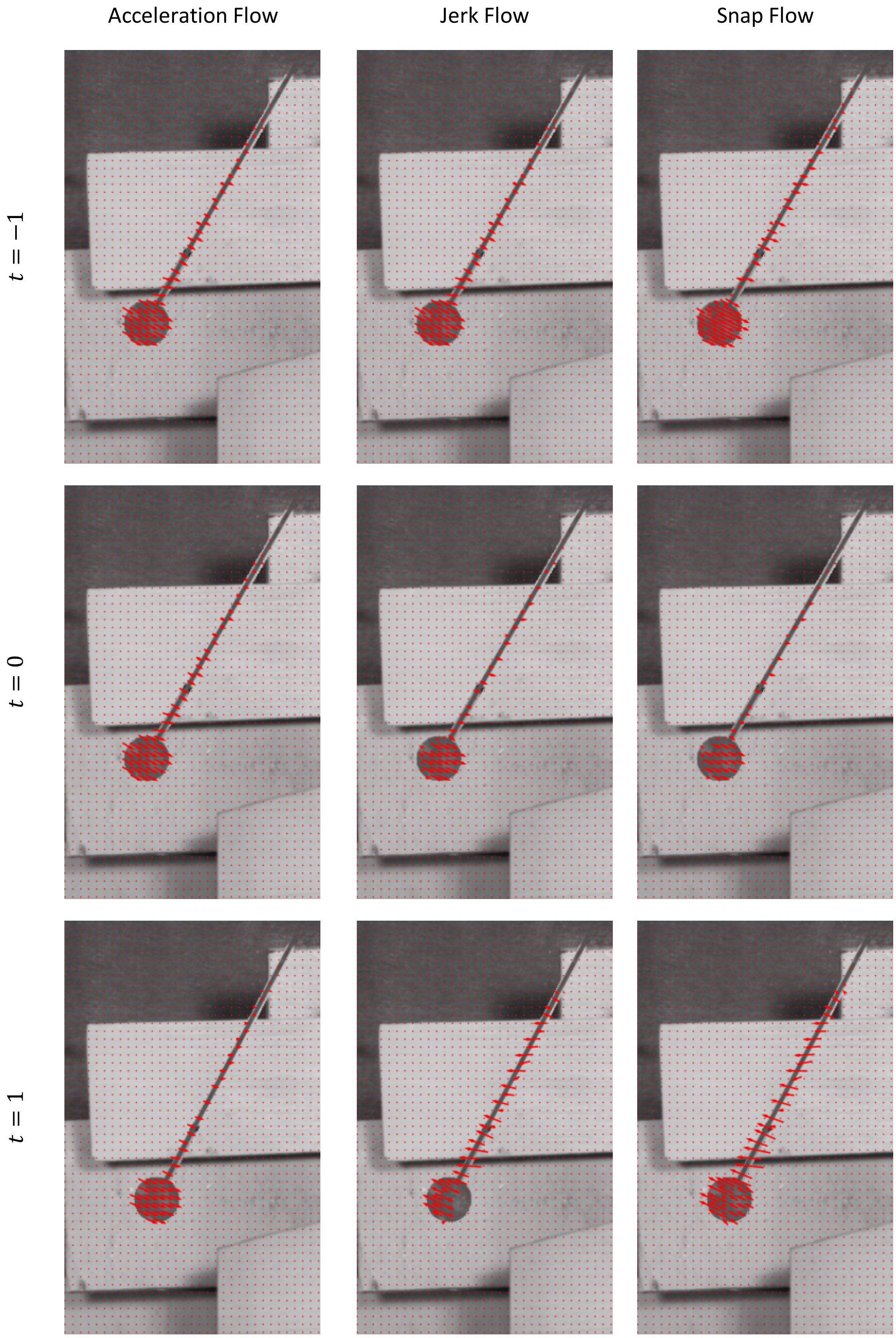
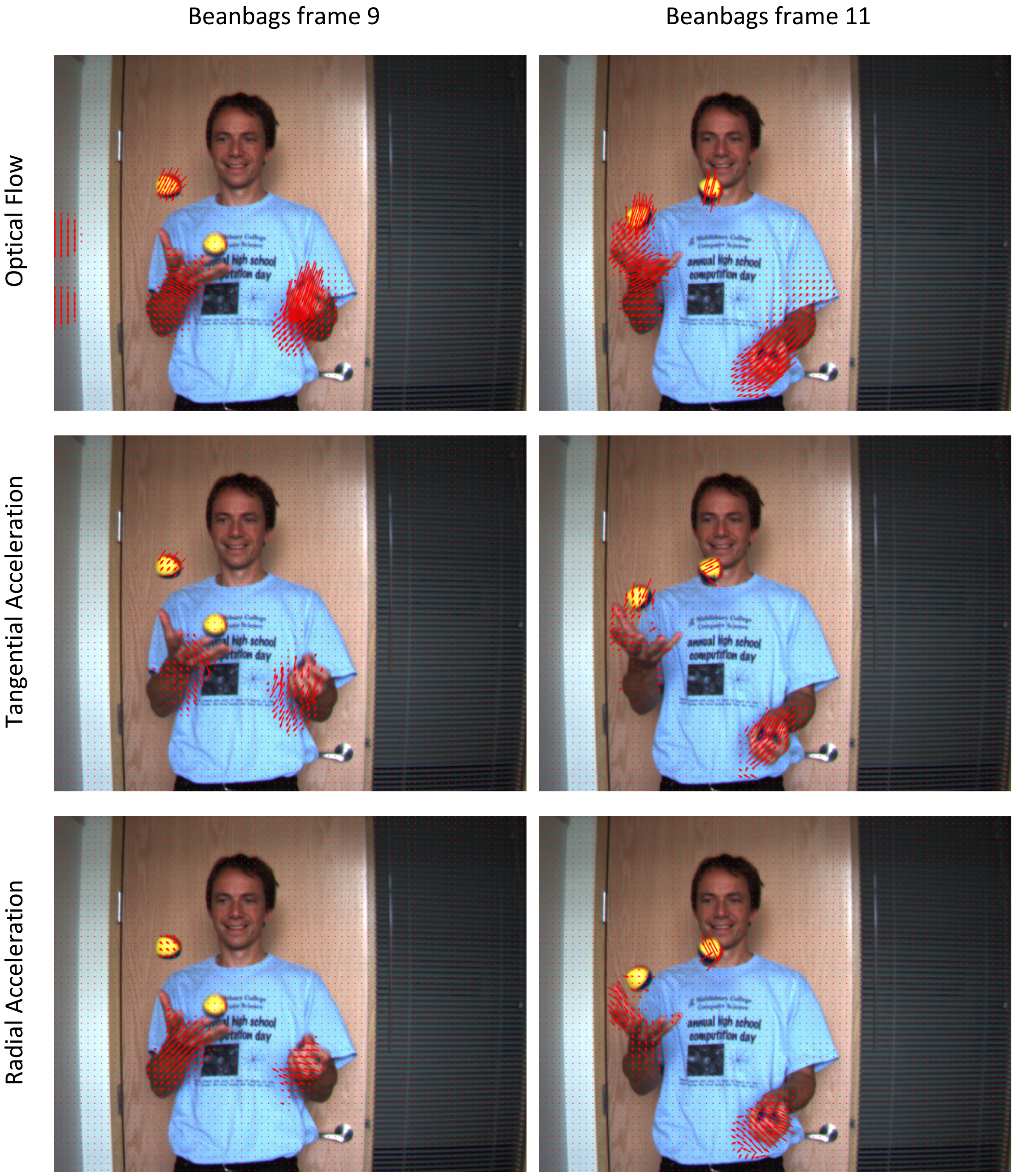


Figure 5.3 Flow fields of synthetic cradle sequence.

In addition to synthetic images, we can disambiguate jerk and snap from real motion fields as well. A few image sequences filmed by high speed camera from Middlebury are selected as the test data for reducing the error caused by low frame rate. We illustrate two sequences in the text, more results can be found in Appendix A and Appendix B.

In Figure 5.4 and Figure 5.5, the results show that the motion fields indeed change with the increasing of order. The constituent parts of snap are too noisy due to the strong constrain (moving along the same arc for five frames), we only illustrate the resultant snap here.



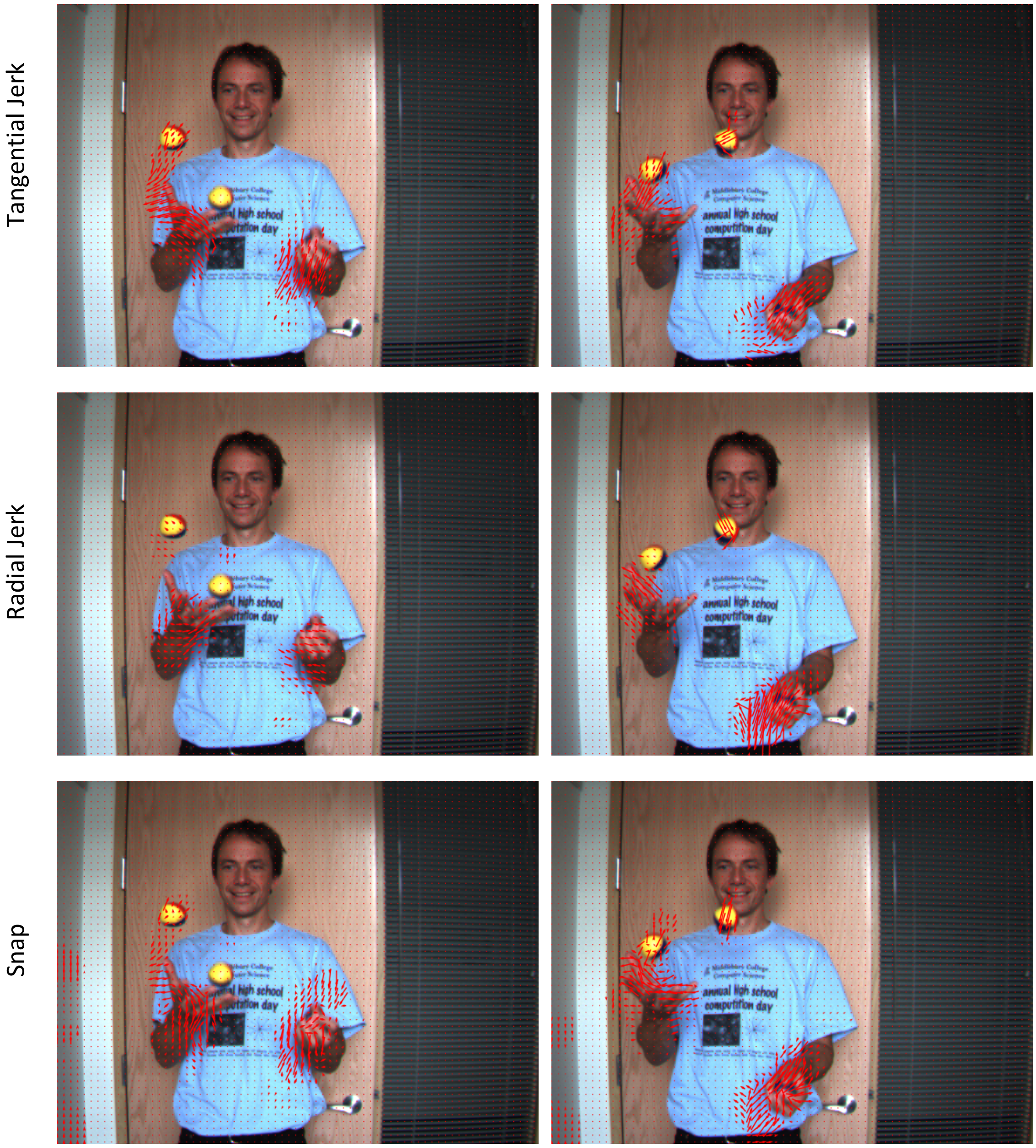


Figure 5.4 The motion fields of Beanbags.

In Beanbags, tangential jerk has larger magnitude than tangential acceleration on the left arm in frame 9 whereas there is barely radial jerk, which means the main change of acceleration is magnitude rather than direction. The right hand shows a similar change. In frame 11, the motion is mostly focused on the right hand and the balls in the air: the large radial jerk flow denotes the acceleration direction of right hand is changing.


DogDance frame 9
DogDance frame 11
Optical Flow
Tangential Acceleration
Radial Acceleration
Tangential Jerk

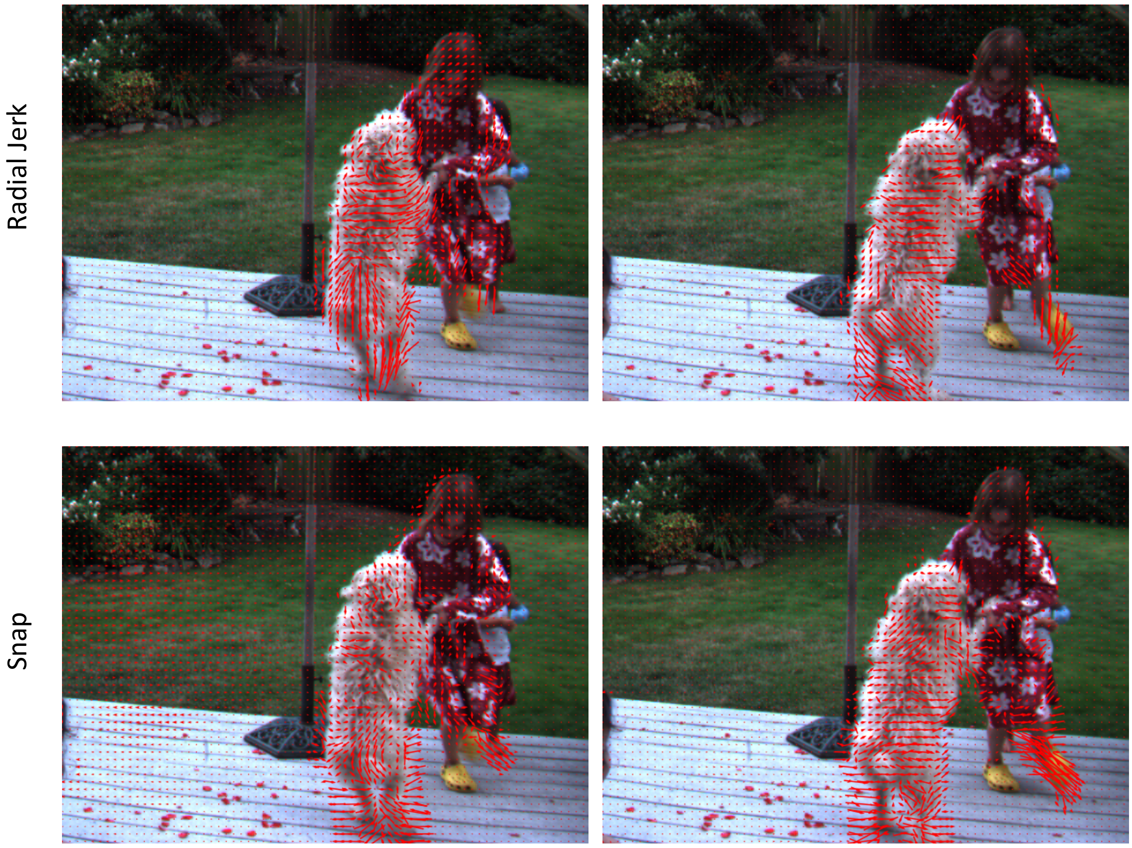


Figure 5.5 The motion fields of DogDance.

In the results of DogDance, the flow fields between different orders are roughly identical, except the dog’s leg in radial jerk and radial acceleration. Again, there is a lot of flow although sometimes they are hard to interpret due to the complexity of real motion.

In kinetics, obtaining acceleration needs three points on the path, four to get jerk and five for snap. With the increasing of motion order, the computation involves more frames since image is a discrete signal. If the difference between frames is finite and relatively large, potentially adding noise to the results. Therefore, in our thesis, the accuracy of motion fields is largely a function of the frame rate and motion intensity. If the frame interval $\rightarrow 0$, we will have the most accurate motion fields. Although the frame rate is currently not high enough for analysing jerk and snap, this preliminary research of disambiguating the motion fields to different components still leans in a step in explaining complex motion in computer vision.

5.4.2 Detecting Heel Strikes via Snap and Jerk

Figure 5.6 shows the normalized radial jerk and snap of a walking cycle sampled every 7 frames. There is considerable amount of flow on the leading foot periodically when the heel strikes on the floor although snap is relatively noisy. Since jerk and snap are higher orders of motion, the flow fields suggest comparatively intense motion. Logically, we wonder whether jerk and snap are able to detect and localize the heel strikes or improve the performance than acceleration.

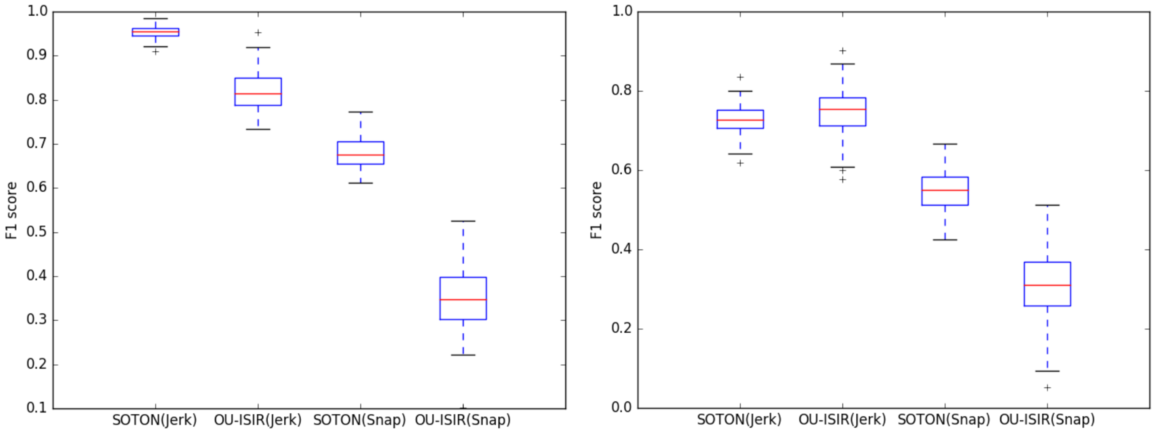
(a) Jerk patterns

(b) Snap patterns

Figure 5.6 There is noticeable amount patterns on a gait cycle.

Figure 5.7 reports the F1 score of detecting heel strike through jerk and snap. The criteria are same with acceleration. As the resolution of all the CASIA data are insufficient and the detection results are very sensitive to the magnitude of motion flow, the flows are only evaluated on SOTON and OU-ISIR.

Jerk shows competitive ability to acceleration on key frame detection and slightly lower on positioning. This is a sign of that jerk can be adapted in gait analysis or other applications on real images. On the other hand, as the highest order of motion snap underperforms acceleration and jerk, furthermore the relatively small ROI in OU-ISIR increases the difficulty.



(a) Key frame detection

(b) Heel positioning

Figure 5.7 F1 score of heel strike detection via jerk and snap.

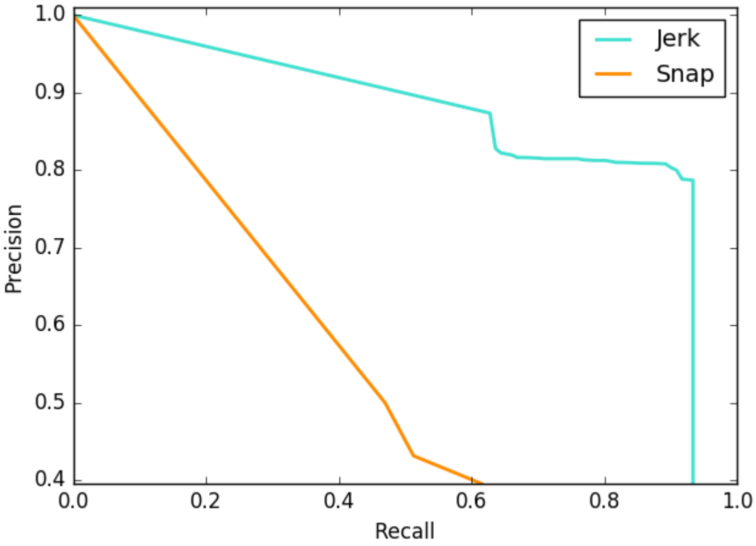


Figure 5.8 Key frame detection PR curves of Jerk and Snap.

The PR curves suggest that jerk has a high area under the curve, which represents it hits both a high recall and a high precision. It performs considerably better than snap on the balance between precision and recall rate.

5.5 Conclusions

In this chapter, motion field is decomposed into its constituent parts. In particular, the notion of acceleration has been extended into detection of jerk and snap (and in their vector format). Analysis of test image sequences and heel strike detection show that the extensions have the power to further discriminate higher orders of acceleration successfully. Clearly the new approaches are ripe for further evaluation and application, perhaps in gait and violence, or in more general image sequence analysis.

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. This could especially be the case when analysing surveillance video, which is often a target of motion-based analysis. In that data there are often problems with low temporal, spatial and brightness resolution, and these can limit the results of a motion-based detection technique. Any development should likely include consideration of the smoothing that is necessary to mitigate limitations inherent in the original data.

Chapter 6 Conclusions and Future Work

6.1 Conclusions

An image sequence compounds many different orders of motion. Previous motion analysis research in computer vision [6], [7] has mostly focused on the displacement between frames, without considering the diversity of motion types. As a more distinct feature than velocity, acceleration clarify certain motions: heel strikes incorporate radial acceleration of the foot whilst walking. Experiments on multiple databases with various conditions show that our method improved reliability on locating of each heel strike within the frame in which it occurs. The immunity of this new approach to the anticipated noise in real CCTV footage is evaluated, radial acceleration is shown to be less sensitive to Gaussian white noise, whilst being effective to changes in viewing angle when compared with the only other known vision-based technique.

This thesis is the first systematic study of classifying motion into different levels in computer vision. There have been a few studies estimating acceleration from optical flow [51], [52]. However there appear to be some limitations in their algorithms and their approaches were not evaluated on benchmark optical flow datasets [7], [85]. First acceleration is derived from the basis of Horn-Schunck but most real motion violates the basic global smoothness assumptions made in the formulation. We sought another way to approximate acceleration fields which is more accurate and able to handle most situations and appears on improvement over Horn and Schunck’s technique on the standard test sequences. Acceleration is decomposed into radial and tangential components, providing a completely new way to understand and disambiguate motions in image sequences. The ability of the new technique is demonstrated by the experiments on a variety of image sequences whereas velocity did not show any obvious difference.

Furthermore, we have explored the higher order motion components: jerk, snap and their constituent parts. Following the idea of acceleration, preliminary results are given to illustrate their different features to motion. Clearly the higher orders of flow are likely to be more sensitive to noise, our research starts a new field for further evaluation and application.

6.2 Future Work

There has been much non-image based research on analysing acceleration and jerk by physical sensors [1], [13], [86], like accelerometer or gyroscopes. The work presented in this thesis makes adapting those researches from physical data to computer images possible. In the future, the following work can be developed towards the following aspects.

6.2.1 Scenes Segmentation

Previous acceleration studies focus on abnormal behaviour detection. Chen and et al. [51] apply acceleration flow to detect abnormal behaviours. Dong and et al. [52] feed optical flow and acceleration flow into a network, both in combination and individually, to detect the violence in scenes. The evaluation illustrates that the most reliable feature is individual acceleration flow. When people fight, their bodies tend to have large acceleration (in many places and with large magnitudes) on their bodies because their arms swing and their feet are kicking. As such, acceleration appears more suited to the detection of rapid changes, consistent with scenes of violence.

We show the acceleration fields with the optical flow on some raw videos from YouTube and Hockey Fight Dataset [87]. In Figure 6.1 there is no fighting and the scenes are mundane; images in Figure 6.2 incorporate various violence scenes. Optical flow fields are illustrated for contrast. In the violence episodes there is considerable detected acceleration and much less focus on irrelevant subjects.

In the tranquil scenes: those scenes which do not contain episodes of violence. 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 data, 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.

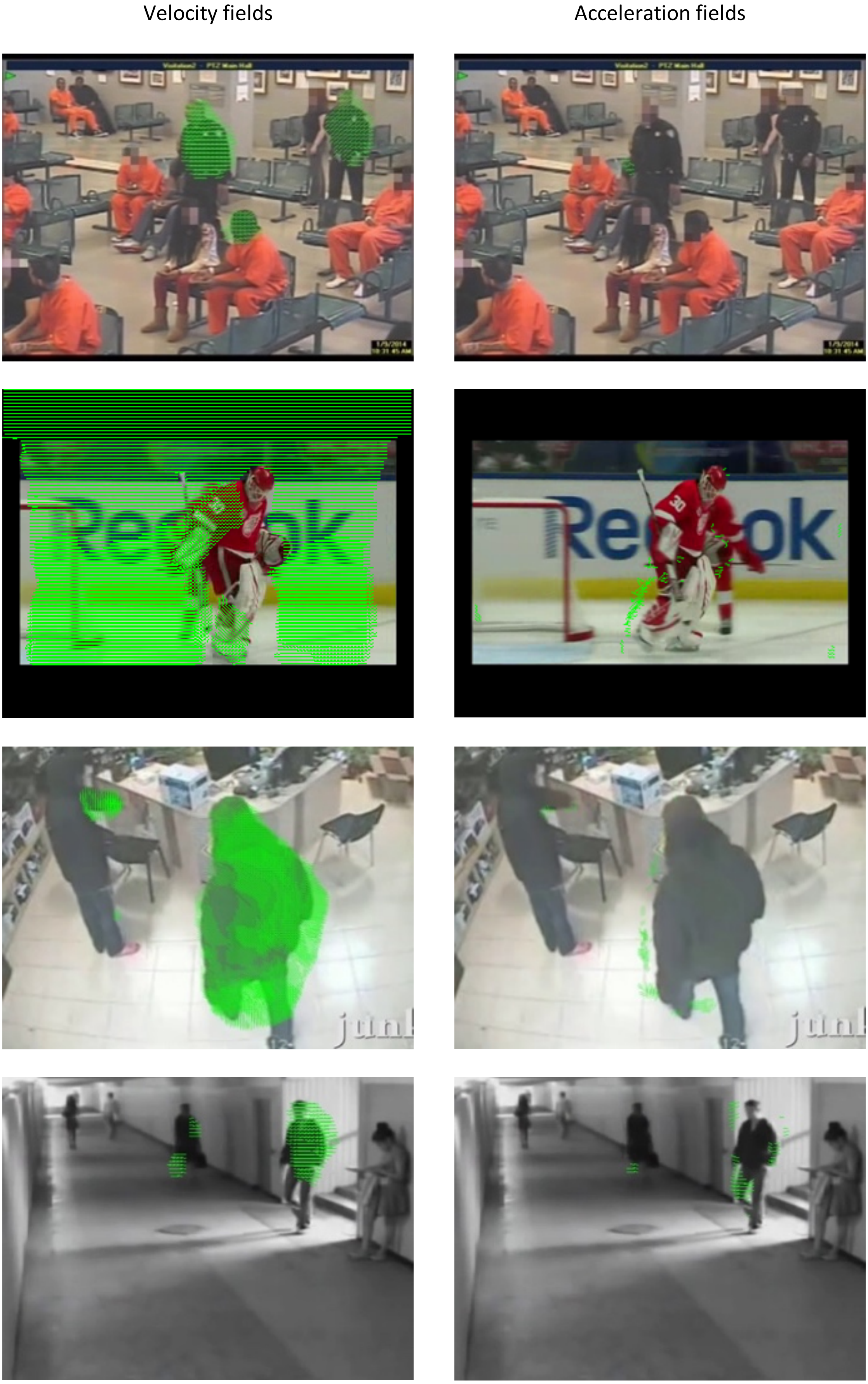
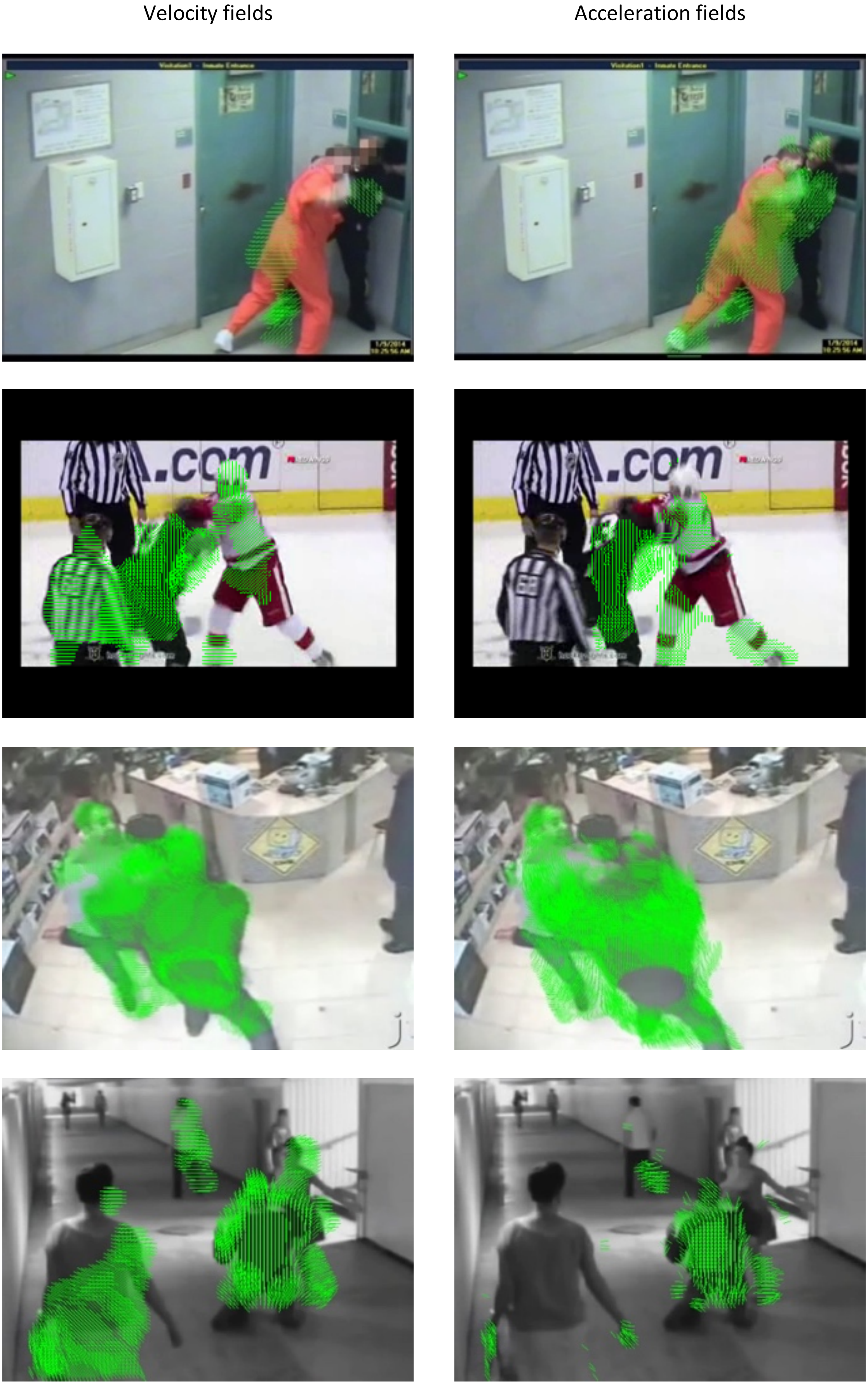

Velocity fields
Acceleration fields

Figure 6.1 The difference between velocity and acceleration within calm scenes.

Figure 6.2 shows the analysis of scenes which contain episodes of violence or criminal behaviours, which are the other parts of the image sequences shown in Figure 6.1. 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. The assault in the third row (which is a continuation of Figure 6.1) 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.


Velocity fields
Acceleration fields

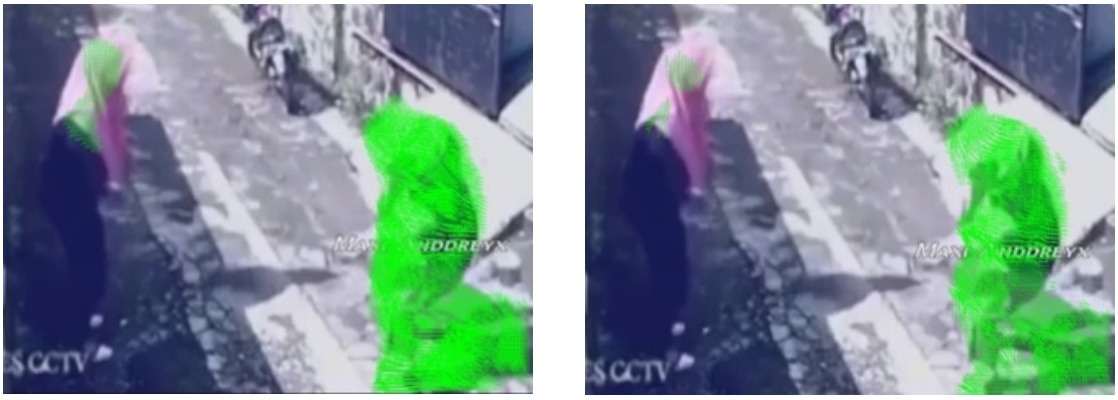
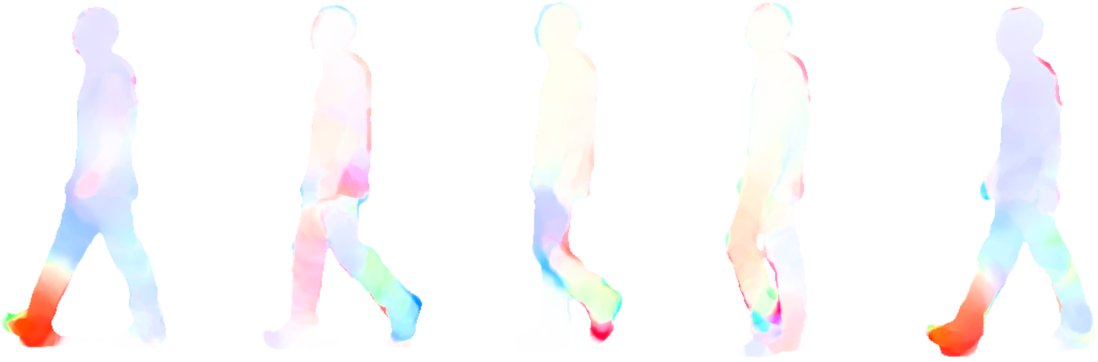


Figure 6.2 The difference between velocity and acceleration within scenes of violence.

6.2.2 Gait Analysis

Accelerometer and gyroscope have been widely used for gait analysis [1], [13] since the movement caused by different gait events is unique. We have proved that radial acceleration and jerk appears periodically on the leading foot at the moment of heel strike in Chapter 4 and Section 5.4.2. Figure 6.3 exhibits the normalized acceleration and jerk magnitude of a whole walking cycle sampled every 5 frames. Acceleration and jerk on each part of the subject are different during different temporal components of the walking cycle.



(a) Acceleration patterns

(b) Jerk patterns

Figure 6.3 The acceleration and jerk patterns on one gait cycle.

This work bridges the gap between analysing gait through motion data from physical sensors to computer images. Previous research of gait analysis has considered velocity and acceleration based on physical data [3], hence we suggest a hypothesis that the multi-orders of flow on each part of the body extracted from computer images can discriminate different phases of walking. Moreover, acceleration can also be used for segmentation. For example, there is rarely substantial acceleration on the upper body, thus presumably the legs and the body can be easily extracted from this observation.

References

[1] P. Connor and A. Ross, “Biometric recognition by gait: A survey of modalities and features,” *Comput. Vis. Image Underst.*, vol. 167, pp. 1–27, 2018.

[2] J. E. Mason, I. Traoré, and I. Woungang, “Gait Biometric Recognition,” in *Machine Learning Techniques for Gait Biometric Recognition*, Cham: Springer International Publishing, 2016, pp. 9–35.

[3] J. Taborri, E. Palermo, S. Rossi, and P. Cappa, “Gait partitioning methods: A systematic review,” *Sensors (Switzerland)*, vol. 16, no. 1. Multidisciplinary Digital Publishing Institute (MDPI), 2016.

[4] 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.

[5] I. Bouchrika and M. S. Nixon, “Gait-based pedestrian detection for automated surveillance,” *Int. Conf. Comput. Vis. Syst.*, no. Icvs, 2007.

[6] D. Fortun, P. Bouthemy, and C. Kervrann, “Optical flow modeling and computation: A survey,” *Comput. Vis. Image Underst.*, vol. 134, pp. 1–21, 2015.

[7] 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.

[8] 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.

[9] M. W. Whittle, *Normal Ranges for Gait Parameters*, Fourth Edi., vol. 3. Harrison Heidi, 2007.

[10] M. S. Nixon, T. Tan, and R. Chellappa, *Human identification based on gait*. Springer, 2005.

[11] P. K. Larsen, E. B. Simonsen, and N. Lynnerup, “Gait analysis in forensic medicine,” *J. Forensic Sci.*, vol. 53, no. 5, pp. 1149–1153, Sep. 2008.

[12] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon, “On Using Gait in Forensic Biometrics,” *J. Forensic Sci.*, vol. 56, no. 4, pp. 882–889, Jul. 2011.

[13] P. B. Shull, W. Jirattigalachote, M. A. Hunt, M. R. Cutkosky, and S. L. Delp, “Quantified self

and human movement: A review on the clinical impact of wearable sensing and feedback for gait analysis and intervention," *Gait Posture*, vol. 40, no. 1, pp. 11–19, May 2014.

[14] 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.

[15] 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.

[16] M. Derlatka, "Modified kNN algorithm for improved recognition accuracy of biometrics system based on gait," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8104 LNCS, Springer, Berlin, Heidelberg, 2013, pp. 59–66.

[17] M. Derlatka and M. Bogdan, "Ensemble kNN classifiers for human gait recognition based on ground reaction forces," in *2015 8th International Conference on Human System Interaction (HSI)*, 2015, pp. 88–93.

[18] C. M. O'Connor, S. K. Thorpe, M. J. O'Malley, and C. L. Vaughan, "Automatic detection of gait events using kinematic data," *Gait Posture*, vol. 25, no. 3, pp. 469–474, 2007.

[19] E. Auvinet, F. Multon, C. E. Aubin, J. Meunier, and M. Raison, "Detection of gait cycles in treadmill walking using a Kinect," *Gait Posture*, vol. 41, no. 2, pp. 722–725, Feb. 2015.

[20] J. Lu, G. Wang, and P. Moulin, "Human identity and gender recognition from gait sequences with arbitrary walking directions," *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 1, pp. 51–61, Jan. 2014.

[21] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A Review of Vision-Based Gait Recognition Methods for Human Identification," in *2010 International Conference on Digital Image Computing: Techniques and Applications*, 2010, pp. 320–327.

[22] A. Świtoński, A. Polański, and K. Wojciechowski, "Human identification based on gait paths," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6915 LNCS, Springer, Berlin, Heidelberg, 2011, pp. 531–542.

[23] 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.

[24] A. F. Bobick and J. W. Davis, “The recognition of human movement using temporal templates,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 3, pp. 257–267, Mar. 2001.

[25] J. Han and B. Bhanu, “Individual recognition using gait energy image,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.

[26] 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.

[27] I. Bouchrika and M. S. Nixon, “Markerless Feature Extraction for Gait Analysis,” in *5th Chapter Conference on Advances in Cybernetic Systems*, 2006, pp. 55–60.

[28] 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.

[29] B. K. B. Horn and B. G. Schunck, “Determining Optical Flow,” *Artif. Intell.*, vol. 17, no. 1–3, pp. 185–203, 1981.

[30] C. Sun, "Fast optical flow using cross correlation and shortest-path techniques." *Image* 29 69, 1999.

[31] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, and H. Seidel. "Complementary optic flow." In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 207-220. Springer, Berlin, Heidelberg, 2009.

[32] M. J. Black and P. Anandan, “The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields,” *Comput. Vis. Image Underst.*, vol. 63, no. 1, pp. 75–104, 1996.

[33] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, “An improved algorithm for TV-L 1 optical flow,” in *Statistical and Geometrical Approaches to Visual Motion Analysis*, vol. 5604, Springer, Berlin, Heidelberg, 2009, pp. 23–45.

[34] 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.

[35] T. Brox, N. Papenberg, and J. Weickert, “High Accuracy Optical Flow Estimation Based on a Theory for Warping,” Springer, 2004.

[36] 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.

[37] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, “SIFT flow: Dense correspondence across different scenes,” 2008.

[38] B. D. Lucas and T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision,” *Imaging*, vol. 130, no. x, pp. 674–679, 1981.

[39] W. Trobin, T. Pock, D. Cremers, and H. Bischof, “An unbiased second-order prior for high-accuracy motion estimation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5096 LNCS, no. 813396, pp. 396–405, 2008.

[40] G. Gordon and E. Milman, “Learning Optical Flow,” *Image (Rochester, N.Y.)*, pp. 83–97, 2006.

[41] A. Dosovitskiy, P. Fischery, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, “FlowNet: Learning optical flow with convolutional networks,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 2758–2766, Apr. 2015.

[42] D. Sun, S. Roth, and M. J. Black, “A quantitative analysis of current practices in optical flow estimation and the principles behind them,” *Int. J. Comput. Vis.*, vol. 106, no. 2, pp. 115–137, 2014.

[43] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, “Performance of optical flow techniques,” *Int. J. Comput. Vis.*, vol. 12, no. 1, pp. 43–77, Feb. 1994.

[44] R. Yaakob, A. Aryanfar, A. A. Halin, and N. Sulaiman, “A Comparison of Different Block Matching Algorithms for Motion Estimation,” *Procedia Technol.*, vol. 11, no. Iceei, pp. 199–205, 2013.

[45] G. Farneb, “Two-Frame Motion Estimation Based on Polynomial Expansion,” *Lect. Notes Comput. Sci.*, vol. 2749, no. 1, pp. 363–370, 2003.

[46] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[47] D. Sun, S. Roth, and M. J. Black, “Secrets of optical flow estimation and their principles,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2432–2439, 2010.

[48] D. Sun, “From Pixels to Layers: Joint Motion Estimation and Segmentation,” *From Pixels to Layers Jt. Motion Estim. Segmentation*, no. May, 2013.

[49] T. Nir, A. M. Bruckstein, and R. Kimmel, “Over-parameterized variational optical flow,” *Int. J.*

Comput. Vis., vol. 76, no. 2, pp. 205–216, 2008.

[50] Y. Sun, J. S. Hare, and M. S. Nixon, “Detecting Acceleration for Gait and Crime Scene Analysis,” in *7th International Conference on Imaging for Crime Detection and Prevention*, pp. 1–6.

[51] 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.

[52] Z. Dong, J. Qin, and Y. Wang, “Multi-stream deep networks for person to person violence detection in videos,” in *Communications in Computer and Information Science*, 2016, vol. 662, pp. 517–531.

[53] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” *Int. J. Rob. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[54] 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, 2004.

[55] 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.

[56] 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. (Cat. No.01CH37164)*, vol. 3, pp. 2299–2306, 2001.

[57] J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, “A survey of video datasets for human action and activity recognition,” *Comput. Vis. Image Underst.*, vol. 117, no. 6, pp. 633–659, 2013.

[58] S. H. Schot, “Jerk: The time rate of change of acceleration,” *Am. J. Phys.*, vol. 46, no. 11, pp. 1090–1094, 1978.

[59] R. T. Collins, R. Gross, and J. Shi, “Silhouette-based human identification from body shape and gait,” in *Proceedings - 5th IEEE International Conference on Automatic Face Gesture Recognition, FGR 2002*, 2002, pp. 366–371.

[60] M. Nixon and A. Aguado, *Feature Extraction and Image Processing for Computer Vision 3rd Ed*. 2012.

[61] L. Wang, T. Tan, H. Ning, and W. Hu, “Silhouette Analysis-Based Gait Recognition for Human Identification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1505–1518, Dec.

2003.

[62] S. Yu, D. Tan, and T. Tan, “A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition,” in *Proceedings - International Conference on Pattern Recognition*, 2006, vol. 4, pp. 441–444.

[63] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, “Robust view transformation model for gait recognition,” in *Proceedings - International Conference on Image Processing, ICIP*, 2011, pp. 2073–2076.

[64] 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.

[65] 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,” *IPSJ Trans. Comput. Vis. Appl.*, vol. 10, no. 1, p. 4, 2018.

[66] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu, “A study on gait-based gender classification,” *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1905–1910, Aug. 2009.

[67] L. Wang, H. Ning, T. Tan, and W. Hu, “Fusion of Static and Dynamic Body Biometrics for Gait Recognition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 2, pp. 149–158, Feb. 2004.

[68] M. Sokolova, N. Japkowicz, and S. Szpakowicz, “Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation,” in *AI 2006: Advances in Artificial Intelligence*, 2006, pp. 1015–1021.

[69] S. H. Schot, “Jerk: The time rate of change of acceleration,” *Am. J. Phys.*, vol. 46, no. 11, pp. 1090–1094, Nov. 1978.

[70] 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.

[71] R. Eichhorn, S. J. Linz, and P. Hänggi, “Transformations of nonlinear dynamical systems to jerky motion and its application to minimal chaotic flows,” *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, vol. 58, no. 6, pp. 7151–7164, Dec. 1998.

[72] P. M. Thompson, "Snap crackle and pop." In *Proc of AIAA Southern California Aerospace Systems and Technology Conference*. 2011.

[73] G. A. M. Meiring and H. C. Myburgh, “A review of intelligent driving style analysis systems

and related artificial intelligence algorithms," *Sensors (Switzerland)*, vol. 15, no. 12, pp. 30653–30682, 2015.

[74] R. Kalsoom and Z. Halim, "Clustering the driving features based on data streams," in *Inmic*, 2013, pp. 89–94.

[75] 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, CIVVS 2009 - Proceedings*, 2009, pp. 23–28.

[76] 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.

[77] B. Bringmann and P. Maglie, "A method for direct evaluation of the dynamic 3D path accuracy of NC machine tools," *CIRP Ann. - Manuf. Technol.*, vol. 58, no. 1, pp. 343–346, Jan. 2009.

[78] H. Nagasaki, "Asymmetric velocity and acceleration profiles of human arm movements," *Exp. Brain Res.*, vol. 74, no. 2, pp. 319–326, 1989.

[79] 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.

[80] 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.

[81] 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.

[82] 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.

[83] 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.

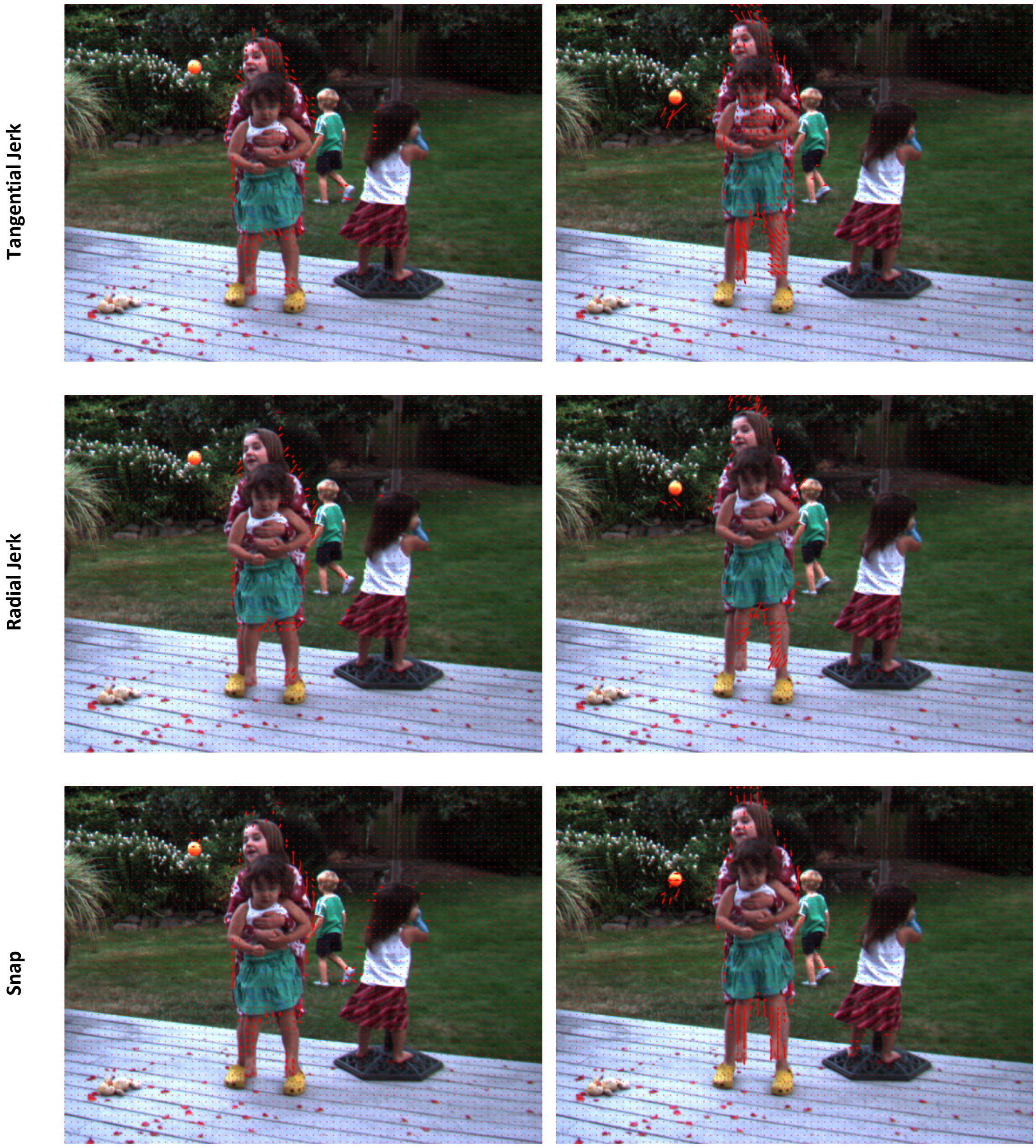
[84] N. Saunier and T. Sayed, "A feature-based tracking algorithm for vehicles in intersections," *Third Can. Conf. Comput. Robot Vision, CRV 2006*, vol. 2006, 2006.

[85] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, “A naturalistic open source movie for optical flow evaluation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7577 LNCS, no. PART 6, pp. 611–625, 2012.

[86] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, “Complex human activity recognition using smartphone and wrist-worn motion sensors,” *Sensors (Switzerland)*, vol. 16, no. 4, p. 426, Mar. 2016.

[87] 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.

Appendix A The Motion Fields of Backyard


Tangential Jerk
Radial Jerk
Snap

Appendix B The Motion Fields of Basketball


Radial Jerk
Snap