Thesis Submitted for the Degree of
Doctor of Philosophy

Non-Locally Regularised Optic-Flow Approaches and Model and Motion based Dense Cell Population Tracking

Sha Yu, B.SE., M.Eng.

Supervisor: Dr. Derek Molloy

Dublin City University
School of Electronic Engineering
April 2015

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____   ID No.:   56125569
Candidate

Date: _____

Acknowledgement

First, I wish to thank Dr. Derek Molloy for his patient supervision, kindness, and support. During the whole course of this Ph.D study, Derek has been supporting all of my interests in the computer vision field, so I have chances to study all the topics that I want to learn.

I also would like to express my sincere gratitude to Dr. Adrian Bors, and Prof Noel O'Connor, who agreed to be my examiners. I enjoyed a very good time talking and discussing with them, during my defense review.

I also wish to thank all current and former members of the Centre for Image Processing and Analysis, for always being so friendly to me. Particulary, I would like to thank Ovidiu, Ketheesan, Tarik, Julius, Julia, Robert, Paul, Laura, Dana, Michele, Sukno, Sean, Aubrey, Brendan, Tony, Ayo, Trish, Min, Victor, and Tian, for their help and encouragement, as well as for the things that I learn from them. Their kindness and friendship will always be cherished.

I would also like to extend my thanks to the staff in the School of Electronic Engineering: Tony Holohan, Patrick, and Breda. A special thank you should go to Robert, Liam, Paul and Billy, for their help to me whenever I went to them.

I also hope to express my sincere gratitude to my teachers XiaoJun Wang and David Molloy, who have provided valuable helps to me, ever since I started my master course study.

I am most grateful to my parents Yin Yinbao and Yu Weisheng, my siblings Yu Chang and Yu Kuai, for their permanent love and support. Without them, I will not have so much happiness during this Ph.D journey as well as in my life.

Contents

List of Figures

[1]http://www.reading.ac.uk/nitricoxide/intro/apoptosis/

List of Tables

Glossary of Acronyms

Acronym	–	Explanation
Acronym	**–**	**Explanation**
2-D	–	Two Dimensional
3-D	–	Three Dimensional
ACM	–	Active Contour Model
DC	–	Data Conservation
DDGVF	–	Dynamic Directional Gradient Vector Flow
f.o.d	–	First Order Derivative
GD	–	Geodesic Distance
GP	–	Geodesic Path
GCV	–	Generalised Cross Valisation
GVF	–	Gradient Vector Flow
GGM	–	Generalised Gaussian Model
GGD	–	Generalised Gaussian Distribution
IC	–	Intervening Contour
IRLS	–	Iteratively Reweighted Least Squares
OF	–	Optic-Flow
OGD	–	Oriented Geodesic Distance
MEP	–	Minimal Energy Path
MSS	–	Motion Supporter Selection
MIR	–	Motion Integration Region
MDP	–	Motion Distribution Prior
MFA	–	Mean Field Approximation
s.o.d	–	Second Order Derivative
TV	–	Total Variation
WT	–	Wavelet Transform

Non-Locally Regularised Optic-Flow Approaches and Model and Motion based Dense Cell Population Tracking

Sha Yu

Abstract

The first half part of this thesis concerns optic-flow (OF) based motion estimation. Recent advances in OF regularisation approaches have emphasised the three important aspects, exploring new motion-integration strategies, looking for improvements or replacements of existing motion spaces, and investigating more suitable motion-distribution priors (MDPs) that can better fit or describe the statistics of a particular motion space. Motivated by that, two motion regularisation approaches, making up the first core contribution of this thesis, have been proposed. First, an oriented geodesic distance based non-local regularisation approach. At the heart of this approach is a novel pairwise-feature-affinity measurement. The new non-local OF approach has been demonstrated particular useful in dealing with two situations: accurately recovering object boundary motion and estimating motion for nearby similar-appearance objects. Experimental results, comparing to leading-edge non-local regularisation schemes, have confirmed the superior performance of the proposed approach. Second, a sparsity&non-sparsity constraint based prior-adaptive regularisation approach, the proposal of which is motivated by that globally fixed MDPs based regularities do not respect local variances of OF statistics. Due to the particular challenge of minimising the involved OF energy functional, a novel Iteratively Reweighted Least Squares (IRLS) and Generalised Cross Validation (GCV) based strategy has also been developed, that can simultaneously optimise the solutions for the flow field as well as the hyperparameter fields involved. Moreover, an exhaustive literature survey on OF regularisation approaches have been provided. This has finally led to a new generalised regularisation formulation, which has been formally clarified, for the first time, in the literature.

The second half part of this thesis focuses on the problem of tracking dense cell populations over phase-contrast image sequences. Quantitative analysis on whole populations of cells, and identification of cell division events plays vital roles in the biomedical research domain. Driven by that, the second major contribution in this work is a model and motion based cell tracking framework. A novel strategy that seamlessly integrates the snake and the OF technique is designed, by enforcing a soft coherence constraint between the model and motion based tracking

techniques. And, the directional gradient vector flow technique is, for the first time, applied to the segmentation and tracking of dense cell populations. The outstanding advantages of the proposed approach are reflected in the following aspects: accurately segmenting ambiguous cell boundaries, correctly tracking partially overlapped cells, consistently tracking elongated cells, and the efficient tracking of large displacement cells. By testing the proposed approach on challenging real cellular datasets, qualitative and quantitative experimental results have indicated that the proposed approach can achieve superior performance, in comparison with the state-of-the-art cell segmentation and tracking approaches. In addition, a third major contribution of this thesis is the development of a motion-occlusion analysis based, automated cell-division detection approach. Through experimentation on different types of cellular datasets, the proposed approach can successfully detect dividing cells with a variety of division behaviour.

Chapter 1

Introduction

The major objective of this chapter is to introduce the motivation for the investigation of regularisation approaches in the optic-flow field, and the development of automated cell tracking systems.

1.1 Overall Motivation

The initial motivation of this work is a cell tracking project. Centered around this topic, there are three frequently investigated techniques, optic flow (Horn and Schunck, 1981), active contour models (ACMs) (Kass et al., 1988), and gradient vector flow (GVF) (Xu and Prince, 1998*b*). Since being introduced, these three techniques have enjoyed great popularity in the computer vision and image processing (CV&IP) community. For example, the significance of optic-flow (OF) algorithms is reflected in many applications, including object tracking/detection, signal compression, 3-D shape or scene reconstruction, correction of camera jitter, image stitching (mosaicing), computer-assisted surgery, special effects, etc. ACMs and their extensions are commonly used in applications like object tracking/recognition, shape modelling, image or volume segmentation, edge detection, path planning, and so on. GVF, from its first introduction as an assisting tool for ACMs, has been extended to other areas, such as image denoising, image restoration, skeletonisation, and object detection.

After a close study, it has been found that, the three techniques are essentially excellent mathematical models that can intuitively allow the encoding of high-level knowledge from human beings. One of the key reasons for their great success owes to the regularisation strategy that has been integrated into these techniques. This has resulted in a great interest in the regularisation approaches involved. More specifically, since the field of regularisation approaches by itself contains a vast amount of literature, it is determined to first start from the regularisation approaches in the OF field. This is due to the consideration that, among many CV&IP algorithms (apart from the ACM and GVF methods), optic flow is one of the fastest growing areas.

It should be clarified that, by investigating into the regularisation approach field, the first purpose of this work is to seek advancement in both OF based motion estimation, and ACM based contour estimation. As another goal, it is expected to also apply the proposed motion or contour estimation approaches to the cell tracking task.

The following section briefly describes the background research into OF estimation and the associated regularisation approaches. Then, the motivation for the research into cell tracking is provided.

- Motivation for the studying of optic flow and the associated regularisation approaches: OF estimation is one of the key problems in computer vision. Estimating the pixel displacements between two images provides a rich source of information that supports a wide variety of analysis tasks. Due to image noise and the well-known aperture problem, the estimation of optic flow is an ill-posed inverse problem. To make the problem feasible, prior knowledge about the solution is required, which is commonly accomplished by means of the so-called regularisation approaches. Because of this, regularisation methods are of crucial importance to accurate OF estimation. Recent advances in regularisation approaches have emphasised three important aspects, namely, exploring new motion integration strategies, looking for improvements or replacements of existing motion spaces, and investigating more suitable motion-distribution priors (MDPs) that can better fit or describe the statistics of a particular motion space. Motivated by this, the first and the third aspects have

been investigated on in this thesis.

- Motivation for developing automated cell segmentation and tracking tools: Cell dynamics is a field of intense current research in which researchers pursue improved comprehension of fundamental processes in cellular and developmental biology. Cell behaviour such as migration (translocation), proliferation (growth and division) and differentiation play a central role in many fundamental biological processes. All of these processes require the orchestrated movement of cells in particular directions to specific locations, and changes of the cellular shape will also often determine the fate of the cell. So, understanding the mechanisms of cell migration and morphology is thus an important goal of biomedical research. The need for automated approaches relies on the fact that the reliable analysis of cellular behaviour involve considerable numbers of cells. In order to prevent influencing cells' lifespan and behaviour, the use of non-fluorescence microscopies, such as phase-contrast microscopy, is usually required. The image recording process can be over a long period of time (usually several days). This will routinely produce thousands of images with low signal-to-noise ratios, which are impractical for manual segmentation and tracking. As a consequence, the development of computer-based techniques that are able to automatically perform cellular segmentation and tracking are highly needed.

1.2 Thesis Outline

Chapter 2 provides the technical background for this research, including the basic theory of OF algorithms, the general formulations of the parametric active contour model and the minimal path model, since the three types of techniques or models are basic components that make up the proposed approaches in this work.

The primary task of Chapter 3 is to present an in-depth survey on existing OF regularisation approaches. The underlying relations between different (sub-)categories of regularisation approaches are established. This leads to a newly generalised formulation of OF regularisation approaches. Based on that, po-

tential research gaps are detected, and new trends that are worthy of future research are also suggested.

In Chapter 4, an oriented geodesic distance based non-local regularisation approach is developed.

The aim of Chapter 5 is two-fold: First, providing a detailed introduction on the research background and challenges about the topic of tracking dense cell populations over phase-contrast image sequences; Secondly, presenting a concentrated literature survey on the most related works in the literature.

Chapter 6 develops a motion-occlusion analysis based, fully automated cell-division detection approach.

In Chapter 7, a novel automated cell tracking framework is proposed, that is particularly designed for two tasks: the segmentation of cell shapes in low contrast phase-contrast images, and the estimation of cell trajectories in highly dense populations. The parametric ACM and the OF technique has been seamlessly combined to realise the established goals.

Chapter 8 presents a sparsity&non-sparsity constraint based prior-adaptive regularisation approach. The chapter also describes the development of an Iteratively Reweighted Least Squares and Generalised Cross Validation based strategy that can simultaneously optimise the solutions for the flow field as well as the hyperparameter fields involved.

Chapter 9 concludes the whole thesis, summarising the contributions and publications that results from this research work, and then suggests promising future directions.

Chapter 2

Technical Background

The main aim of this chapter is to provide the necessary technical background for this research. Section 1 firstly introduces the basic theory of optic-flow (OF) algorithms, and then explains two classic OF techniques. For an overall understanding of the research trend in the OF domain, special emphasis has been given to the topics, from challenges of constructing proper models, to strategies for optimising the solutions. Section 2 describes the general formulations of two types of deformable contour models, namely the parametric active contour model and the minimal path model, since they will be important components that consist of the contributions of this research. To facilitate a better understanding of the deformable models involved in this research, representative extensions and related strategies regarding the metric design are investigated. The last part of Section 2 provides a comprehensive explanation on the central connections between different families of deformable contour models of interest.

2.1   Optic Flow Algorithms

The goal of OF estimation is to compute an approximate motion field for time-varying image sequences. All OF approaches rely on the temporal conservation of some image properties, e.g. conservation of pixel intensity or gradient values. Commonly used elements are pixels, since they can be easily extracted and

lead to dense measurements. In addition to data conservation, spatial motion coherence is another essential assumption within traditional OF methods. To provide an introduction on how standard OF algorithms are structured, the following sections first present short overviews for the two assumptions that underpin the OF theory.

2.1.1   Data Conservation Assumption

The most straightforward data conservation (DC) assumption is the Intensity Constancy Assumption (ICA), which assumes that pixel intensities are translated from one image frame I_1 to the next I_2 (Fleet and Weiss, 2005),

$$I_2(x + u, y + v) = I_1(x, y) \tag{2.1}$$

where $I(x, y)$ is image intensity as a function of space point $\mathbf{x} = (x, y)^T$, and $\mathbf{v} = (u, v)^T$ is the 2-D velocity field of image pixels.

Although this constancy assumption works fine in many cases, algorithms that rely on this prerequisite cannot deal with images with local or global changes in illumination. For this reason, other DC assumptions that are invariant to brightness changes are proposed. These assumptions have the following formulations (Brox et al., 2004, Papenberg et al., 2006):

- Constancy of the gradient: A global change in illumination will shift and/or scale gray values of the associated images, while not affecting the gradient. The spatial gradients between two images can be considered as constant during motion: $\nabla I_2(\mathbf{x} + \mathbf{v}) = \nabla I_1(\mathbf{x})$.

- Constancy of the Hessian: Notice that the gradient-based conservation formulation inherently assumes the constancy of first-order image derivatives. It is not difficult to also consider higher-order derivatives for the formulation of constancy assumptions. One choice for including second-order derivatives is the Hessian matrix $\mathcal{H}_2$: $\mathcal{H}_2 I_2(\mathbf{x} + \mathbf{v}) = \mathcal{H}_2 I_1(\mathbf{x})$.

- The Laplacian, i.e. the trace of the Hessian, can also be used to formulate a constancy assumption: $\Delta I_2(\mathbf{x} + \mathbf{v}) = \Delta I_1(\mathbf{x})$, where $\Delta = \partial_{xx} + \partial_{yy}$ denotes the spatial Laplacian.

Table 2.1: Overview of some representative constancy assumptions, and their formulations.

Constancy assumption	Data term formulation
Gray value	$\min\limits_{\mathbf{v}} \int_\Omega |I_2(\mathbf{x} + \mathbf{v}) - I_1(\mathbf{x})|^2$
Gradient	$\min\limits_{\mathbf{v}} \int_\Omega |\nabla I_2(\mathbf{x} + \mathbf{v}) - \nabla I_1(\mathbf{x})|^2$
Hessian	$\min\limits_{\mathbf{v}} \int_\Omega |\mathcal{H}_2 I_2(\mathbf{x} + \mathbf{v}) - \mathcal{H}_2 I_1(\mathbf{x})|^2$
Laplacian	$\min\limits_{\mathbf{v}} \int_\Omega |\Delta I_2(\mathbf{x} + \mathbf{v}) - \Delta I_1(\mathbf{x})|^2$

Considering that the gradient and the Hessian contain directional information, they can be employed in estimating translocation and divergent motions. However, in the case of estimating rotation motion, the scale-value based constancy assumptions, such as the gray value and the Laplacian, are more suitable.

A short summary of the data terms that are based on the previous constancy assumptions is given in Table 2.1.

2.1.2   Spatial Motion Coherence Assumption

The 2-D velocity $\mathbf{v}$ cannot be recovered with two unknowns, u and v, from only the identity function as defined in Equation (2.1). To deal with this under-constrained problem, the motion smoothness constraint is traditionally used, by assuming that neighbouring pixels share the same 2-D velocity. Two classic regularisation schemes are described here:

- The Horn-Schunck (HS) scheme (Horn and Schunck, 1981):

$$E_{Smooth} = \sum_{\mathbf{x} \in \Omega} \sum_{\mathbf{x}' \in \mathcal{N}_{\mathbf{x}}^1} (\mathbf{v}_{\mathbf{x}} - \mathbf{v}_{\mathbf{x}'})^2 \qquad (2.2)$$

  where $\mathcal{N}_{\mathbf{x}}^1$ denotes the local neighbourhood centered at an arbitrary pixel $\mathbf{x}$. According to the formulation, for each pair of directly adjacent pixels, the pairwise motion similarity is calculated. By summing up the local motion deviations for all pixels inside an image domain Ω, the smoothness term measures the implicit, global motion coherence for the

flow field $\mathbf{v}$. A combined formulation with a gray-value based data term and the regularisation term can be written as:

$$E(\mathbf{v}) = \sum_{\mathbf{x} \in \Omega} (I_2(\mathbf{x} + \mathbf{v}) - I_1(\mathbf{x}))^2 + \alpha E_{Smooth} \qquad (2.3)$$

where $\alpha > 0$ is usually a constant parameter that controls the balance between the smoothness and the DC term. The minimisation of the convex energy function (Equation (2.3)) is usually achieved by Euler-Lagrange equations (Brox et al., 2004), leading to a dense OF field $\mathbf{v}$.

- The Lucas-Kanade (LK) scheme (Lucas and Kanade, 1981):

$$E(\mathbf{v_x}) = \sum_{\mathbf{x}' \in \mathcal{N}_{\mathbf{x}}^2} \omega_{\mathbf{x},\mathbf{x}'} (I_2(\mathbf{x}' + \mathbf{v_x}) - I_1(\mathbf{x}'))^2 \qquad (2.4)$$

The LK method assumes that the unknown OF vector is constant or similar within a local or slightly larger neighborhood $\mathcal{N}_{\mathbf{x}}^2$ that is centered at a pixel $\mathbf{x}$. The weight value $\omega_{\mathbf{x},\mathbf{x}'} > 0$ determines how much the motion of the centered pixel $\mathbf{x}$ should be similar to its neighbour $\mathbf{x}'$. A common way to calculate $\mathbf{v_x}$ is to minimise $E(\mathbf{v_x})$ by Least-Squares Estimation. Considering that LK based OF approaches produce sparse flow fields, a post-processing step for flow interpolation can be made to obtain dense fields. Notice that the smoothness assumption and the DC assumption for a pixel $\mathbf{x}$ is simultaneously encoded in the single energy term $E(\mathbf{v_x})$.

Explicit and Implicit Regularisation based Approaches

According to the definition of the smoothness term in Equation (2.2), the HS algorithm regularity explicitly imposes motion coherence between adjacent pixels, by calculating the deviation between the motion vectors involved. The under-constraint problem associated with each single motion vector is solved by jointly estimating all of the motion vectors together.

Within the LK OF scheme, a unified energy term is formulated as Equation (4.2), and OF vectors are estimated by matching similar-appearance patches between images. For each patch-to-patch matching, a motion vector is estimated that is expected to make the value of $I_2(\mathbf{x}' + \mathbf{v_x}) - I_1(\mathbf{x}')$, for any involved $\mathbf{x}' \in \mathcal{N}_{\mathbf{x}}^2$, to be as small as possible. During a single process of patch-to-patch

mapping, the LK type of formulation implicitly assumes that pixels within the considered neighbourhood move coherently together.

The HS style of OF algorithms naturally result in dense vector fields, and the LK type of estimation usually produces sparse vector fields. To distinguish these two types of frameworks in this thesis, they are respectively referred to as the global-energy type OF, and the local-energy type OF.

2.1.3 Challenges of Optic Flow Estimation

Problems Associated with Data Matching

Data matching can be difficult under circumstances of illumination changes, textureless regions, clustering of similar-appearance objects, non-rigid objects with appearance changes between frames, etc. By imposing the DC constraint, a feature matching process is inherently conducted within the OF estimation. For this reason, sophisticated feature descriptors and matching processes that are based on SIFT (Scale-Invariant Feature Transform) (Lowe, 1999), HOG (Histogram of Oriented Gradients) (Dalal and Triggs, 2005), or the nearest neighbour field (NNF) (Barnes et al., 2009) techniques have been recently considered in the literature. For example, Brox and Malik (2011) introduce an extra data-constancy term that matches SIFT-based feature descriptors between image pairs. In (Xu et al., 2012), improved motion estimation results have been reported by adaptively fusing SIFT-based feature matching into traditional OF frameworks. Most recently, NNFs (Barnes et al., 2009) have been investigated in assisting OF estimation (see the work of Chen et al. (2013)), and have achieved superior motion estimation results (see the reported results on the well-known Middlebury Benchmark[1]). In theory, advanced feature descriptors include richer data information, and can thus help to improve the motion estimation. The main issue, which is unavoidable in either traditional or recently proposed data matching schemes, involves feature matching for regions with weak textures, where reliable key features are insufficient or absent. So, the correspondences of textureless regions remain ambiguous. The other

---

[1]http://vision.middlebury.edu/flow/data/

important issue is the considerably increased computation cost, compared with using relatively simple DC terms.

Problems Associated with Regularisation

In order to solve the ill-posed problem of OF estimation, prior knowledge is required on typical motion fields to be restored. The regularisation theory assumes that the underlying motion field has some spatial smoothness (the traditional prior) or global sparsity[2] (a recently suggested prior, see the works of Shen and Wu (2010), Jia et al. (2011)). Considering that motion fields can have varying motion statistics across different images or different patches in a single image, a natural problem is how to select suitable regularities for a targeted motion field. There is a large amount of literature on the topic, with much on-going research, e.g., designing more advanced motion-smoothness priors based on higher-order models (Lee et al., 2010, Yuan et al., 2007, Onkarappa and Sappa, 2013, Kwon et al., 2010), seeking sparsity priors that can more compactly describe the structured motion information (Shen and Wu, 2010, Jia et al., 2011), attempting robust segmentation strategies to facilitate the motion integration requirement (Sun, Sudderth and Black, 2010, Sun et al., 2012, 2013), reducing the computation cost that comes with the introduced complex prior models (Krähenbühl and Koltun, 2012), etc. Details with regard to the topic of motion regularisation will be included in Chapter 3.

The Motion Occlusion Problem

In OF estimation, a motion occlusion can be caused by an occluding surface moving in front of an occluded surface, or due to camera motions. An occlusion point can thus be associated to the photometric value that perceptually "appears" or "disappears" between two consecutive frames. This phenomenon violates the DC assumption, and thus poses difficulties in motion estimation. In some approaches, occlusions are detected separately from the process of

---

[2]In this work, a global sparsity means that a sparse gradient prior is assumed on the whole motion field.

OF estimation. For example, after estimating a forward OF field and a backward field for two consecutive images, the mutual consistence between two OF fields is used to determine an occlusion map (Baker et al., 2011, Xiao et al., 2006). Other approaches try to jointly estimate disparity and occlusion, by taking occlusions as important cues for depth and motion discontinuities (Sun, Sudderth and Black, 2010, Unger et al., 2012). Still, other approaches can be referred to (Ayvaci et al., 2010, Ince and Konrad, 2008, Xiao et al., 2006).

The Energy Minimisation Problem

Due to the fact that non-convex energy functionals are involved in modern OF algorithms, special strategies or treatments are usually needed to minimise the associated functionals. For convex energy functionals, techniques such as gradient descent (Rudin et al., 1992a), dual formulations (Chan et al., 1999), split Bregman (Goldstein and Osher, 2009) schemes are effective. While, when problems are non-convex, these techniques usually fail because they get "stuck" at local minima. In (Brox et al., 2004), since the proposed energy functional is nonlinear, with respect to both the data and smoothness terms, nested loops of fixed-point-iteration strategies are used. In each loop, linearisation treatments aim to remove the nonlinearity of the associated energy functional. After that, common numerical methods are employed to solve the previously resulting linear system (Brox et al., 2004). Black and Anandan (1996), Sun, Roth and Black (2010) both exploit the Graduated Non-Convexity (GNC) method. Starting from constructing a corresponding convex approximation for the original non-convex functional, the GNC method initially solves the simplified problem, and progressively transforms that problem until it is equivalent to the difficult, original non-convex minimisation problem. In (Xu et al., 2012), the mean field approximation (MFA) (Geiger and Girosi, 1991) enables solving the objective functional, which involves both discrete and continuous variables. Considering that there is no recent, focused survey on this topic, interested readers can be referred to still other works such as (Lempitsky et al., 2008, Steinbrucker et al., 2009, Sun et al., 2012, Li et al., 2013), to find more relevant treatments. In conclusion, accurate estimation of optic flow is still a challenging task that usually requires addressing difficult energy minimisation

problems. It is believed that more efficient as well as general techniques are required.

Reducing the Computational Cost

Recently, due to the introduction of non-local regularisation methods (Sun, Roth and Black, 2010, Werlberger et al., 2010), and high-order derivatives based prior models (Roth and Black, 2007, Baker et al., 2011) into the OF field, the minimisation of the objective functionals has become extremely costly, because computation time scales with the n-th power of the number of involved pixels in terms of the non-local or high-order models. To reduce the computation demand, some researchers resort to approximation formulations or efficient strategies for implementation. For example, in (Sun, Roth and Black, 2010), a weighted median filtering based scheme (Li and Osher, 2009) tries to speed up the implementation. Krähenbühl and Koltun (2012) employ the high-dimensional Gaussian filtering operation, which can be performed in linear time, so as to make the implementation efficient. In (Kadri-Harouna et al., 2013), the difficult problem of directly calculating high-order derivatives is circumvented within a wavelet based OF implementation, so a low computational complexity is accomplished. Alternatively, Graphics Processing Unit (GPU) accelerated implementations have been applied, e.g., see the works of Werlberger et al. (2009), Ranftl et al. (2012).

2.2 Energy-Minimal Contour Models

Since the original work on active contour models by Kass et al. (1988), extensive research has been performed on deformable contours. The active contour model and its extensions are examples of general techniques that match a deformable model to a portion of image features, using the energy minimisation strategy. In this research work, active contours and some extended deformable contours are uniformly called energy-minimal contours. Specifically, two important types of energy-minimal contour models, namely the parametric active contour model and the minimal path model, are introduced, considering that

they play important roles in this research.

2.2.1 Parametric Active Contour Models (ACMs)

Parametric active contours, or snakes, were originally introduced by Kass et al. (1988), for solving various computer vision problems, such as edge/contour detection, object segmentation and tracking. A snake is an energy minimising curve, with a predefined energy equation that is structured in a way to achieve a desired deformation. Under the influences of different energies (as defined in (Kass et al., 1988)), a snake is able to change shape and position within the image; the snake is then expected to have converged onto an object of interest, when the minimum of the total energy is reached.

General Formulation

Mathematically, a snake is defined as a parametric contour, represented by a close or open curve $C(s) = (x(s), y(s))$, which is parameterised by s in the range of $[0, 1]$. In practice, the curve is represented by a number of discrete control points that are usually referred to as control points, or snaxels. The total energy of the snake, $E(C)$, is given by the functional

$$E(C) = \int_1^0 (\alpha |C_s|^2 + \beta |C_{ss}|^2 + E_{ext}(C(s)) + E_{con}(C(s)))ds, \qquad (2.5)$$

where C_s and C_{ss} denote partial derivatives of C with respect to s, and respectively models the contour's elasticity and smoothness properties. $|\cdot|$ denotes total variation. The associated coefficients α and β are pre-determined weighting parameters that are used to balance the effects of the first two terms in Equation (2.5), which are jointly called the internal energy of the snake. The third term is the snake's external energy, with E_{ext} formulated in different ways. For example, E_{ext} can be defined as $E_{ext}^1 = -G_\sigma * I$, or $E_{ext}^2 = -|\nabla(G_\sigma * I)|^2$, where I is the given grey image. The operation of $G_\sigma * I$ smoothes the original image by convolution using a Gaussian filter. The symbol ∇ in the equation E_{ext}^2 represents the gradient operation. E_{con} gives rise to some additional constraints, such as repulsion or attraction constraints (see demonstrations in

(Kass et al., 1988)). In practice, the fourth term E_{con} is optional with respect to specific applications.

The minimisation of the internal energy imposes a certain tension and rigidity on the curve, helping the snake to overcome image noise and have some preferred shape. In the case of Equation (2.5), a round shape is preferred, since the snake always tends to be short and smooth. The minimisation of the third term drives the curve towards salient image features such as light/dark regions, edges, etc. For example, when the function E_{ext}^1 is chosen, the snake will have a lower energy when locating on brighter regions. The effect of the function E_{ext}^2 makes the snake move towards high gradient regions, i.e., normally object edges.

A snake that minimises $E(C)$ (with E_{con} not included) must satisfy the Euler equation :

$$\alpha C_{ss} + \beta C_{ssss} + \nabla E_{ext} = 0. \tag{2.6}$$

where C_{ss} and C_{ssss} respectively denote the second and fourth partial derivatives of C with respect to s. α and β are the same coefficients as defined in Equation 2.5. The three terms can be viewed in order as the elastic force $F_{elastic}$, the rigid force F_{rigid}, and the external force F_{ext} of the snake. The snake will change shape and position because of the competition of the three forces, and will reach equilibrium when the forces are balanced by each other.

Representative Extensions

Two major problems with the snake model are its sensitivity to noise and the short capture region. For example, if the snake is initialised not close enough to the features of interest, the snake can become stuck in a local minima. In order to deal with the limitations, both the internal and external energies of snakes have been adjusted or modified in the literature, leading to enhanced overall performance. Particularly, efforts have been made to improve the modelling of the appearance (edge, shape, colour, etc) properties of targeted objects, to accommodate particular segmentation or tracking tasks. In this regard, some representative extensions of the original snake model are introduced as following.

Ray et al. (2002) proposed to impose shape and size constraints, in forms of new internal energies, into the ACM, for tracking rolling leukocytes. Charmi et al. (2008) integrated the geometric shape-prior information into parametric active contours. The shape-prior constraint is formulated as an extra internal-energy term, encouraging the snake to have similar Fourier-shape-descriptor coefficients with the corresponding coefficients of a template shape. The method enhances the model's robustness to noise and occlusion, because the evolving snake is always guided to a shape similar to the given template/prototype. In (Srikrishnan and Chaudhuri, 2011), a so-called distraction-free ACM is proposed. The main contribution is a new internal force that dynamically varies along the snake curve during its evolution, to encourage the snake to locate on a continual edge. This is achieved by penalising snake segments lying on nearby distracting edges from local clutters. The physical intuition behind the method is that, an object edge could consist of weak and strong segments but should be generally continuous.

The gradient vector flow (GVF), proposed by Xu and Prince (1998b), represents a new external force field that is created by spreading edge information of an image farther away, via diffusing image gradient vectors. The advantage of the GVF technique is reflected in the increased capture region for snakes. In (Ray and Acton, 2004), a motion gradient vector flow (MGVF) is proposed, by extending the GVF technique to take into account the prior directional information of the target movement. The constructed new external energy encourages an active contour to evolve towards the prior direction. By this means, the MGVF based active contour is able to track a relatively fast-moving object. Cheng and Foo (2006) propose a new type of external force for snakes, named dynamic directional gradient vector flow. The main idea of the approach exploits the image gradients in both x and y directions, and deals with external force fields with respect to the two directions separately. In snake deformation, the DDGVF field is utilised dynamically, according to the orientation of the snake in each iteration. By this means, the DDGVF snake provides reliable segmentation in situations of clustered objects (which pose confusion for segmentation). Wang et al. (2009) proposed a parametric ACM using a new concept, called fluid vector flow (FVF), that simulates the fluid flowing along object boundary, and generates external force fields dynamically

to drive the contour evolution. FVF snakes have shown large capture ranges, and are able to extract acute concave shapes, due to its non-static external force fields. The optic flow active contour, proposed by Hamou and El-Sakka (2010), is a technique that combines the pre-calculated OF magnitude and the image-edge map together for calculating a GVF as an external force field for snakes. The OF field thus provides additional information to the ACM, and helps to improve the performance of snakes in object tracking. Other external forces for snakes include: the "balloon force" (Cohen, 1991), the curvature vector flow (Gil and Radeva, 2003), the vector field convolution (Li and Acton, 2007) technique, the generalised GVF (Xu and Prince, 1998a), and the edge-preserving GVF (Li et al., 2005) based forces.

2.2.2 Minimal Path Model (MPM), Geodesics

Definition of Geodesic Path, Geodesic Distance

In the minimal path technique that is originally proposed by Cohen and Kimmel (1997), an energy E of the form

$$E(\gamma) = \int_\gamma W(\gamma(s))|\gamma'(s)|ds \tag{2.7}$$

is computed along the curve γ within the image domain Ω. $\gamma'(s)$ denotes the derivative of γ, so $\int_\gamma |\gamma'(s)|ds$ measures the Euclidean length of the curve. A potential $W : \Omega \to \mathbb{R}^+$ is usually built from a given image $I : \Omega \to \mathbb{R}^+$, where $W > 0$, that takes lower values near the desired features of the image I. Given a metric function $W(\cdot)$, the goal is to extract a curve that minimises the energy functional $E : \Gamma_{x_1,x_2} \to \mathbb{R}^+$ between two user-defined points x_1, x_2:,

$$E(\gamma) = \min_{\gamma \in \Gamma_{x_1,x_2}} \int_\gamma W(\gamma(s))|\gamma'(s)|ds \tag{2.8}$$

where Γ_{x_1,x_2} is the set of all paths connecting x_1 to x_2, and s is the arc length parameter. The curve connecting x_1 to x_2 that globally minimises the energy $E(\gamma)$ is called the *minimal path* or the *geodesic path* between x_1 and x_2. The geodesic path is represented by C_{gd}. Additionally, the minimal energy associated with the geodesic path is referred to as the *geodesic distance*, denoted as $D_{gd}(x_1, x_2)$, between the two pre-defined endpoints.

A note on implementation: The solution of this minimisation problem is obtained through the computation of the geodesic distance map (GDM) $U : \Omega \to \mathbb{R}^+$ associated with x_1. The GDM for the source point x_1 is defined as the minimal energy that is integrated along a path between x_1 and any point x of the domain Ω:

$$\forall x \in \Omega, U^{x_1}(x) = D_{gd}(x_1, x) = \min_{\gamma \in \Gamma_{x_1,x}} \int_\gamma W(\gamma(s))|\gamma'(s)|ds. \tag{2.9}$$

Here, the values of U^{x_1} can be considered as the arrival times of a front propagating from the source x_1 with velocity $\frac{1}{W}$ and $U^{x_1}(0) = 0$. U^{x_1} satisfies the Eikonal equation (Cohen and Kimmel, 1997, Peyré et al., 2010, Benmansour and Cohen, 2011):

$$|\nabla U(x)| = W(x). \tag{2.10}$$

This equation can be solved, for example by the Fast Marching algorithm[3], numerically in $O(Nlog(N))$ operations on a discrete, isotropic grid of N points.

Metric Design for MPMs

In practice, the most important task associated with the MPM is to design a metric W, in order to have meaningful geodesics. The choice of potential W depends on the application. Here are some examples of possible choices, for the processing of an input image I:

Pixel-value based metric:

$$W(x) = |I(x_1) - I(x)| \tag{2.11}$$

where x_1 denotes the source pixel. This metric $W(x)$ is defined to be low at pixels whose value is close to the intensity of the source pixel x_1.

Intensity based metrics:

$$W(x) = I(x) \tag{2.12}$$

For the definition, W has a low value where the pixel intensity is small. So, for two given endpoints x_1, x_2, the associated geodesic path will try to include

---

[3]The Fast Marching algorithm, introduced by (Sethian, 1995), is a numerical algorithm that is able to catch the viscosity solution of the Eikonal equation of the form $\|\nabla U\| = W$.

dark pixels as components, rather than bright ones. Since in most medical images, vessels can be made darker than the background, by choosing this type of metric, geodesic paths can be used to extract dark contours of the image (Peyré et al., 2010, Benmansour and Cohen, 2011), i.e. vessels. In other applications, such as road tracking in vehicles, this potential function can be also applied.

Edge based metric (Peyré et al., 2010):

$$W(x) = \frac{1}{\epsilon + G_\sigma * |\nabla I(x)|} \tag{2.13}$$

where G_σ is a Gaussian kernel of variance σ^2, and the symbol $*$ denotes convolution. In several applications, the curve of interest is located near areas of large variation of intensity (i.e. high gradient) in the image. Taking advantage of the edge based metric, the geodesic path is possible to extract the object boundary, since it always tries to follow high gradient pixels.

MPMs based on the previous metrics are isotropic cases, i.e. the metrics only exploit scalar values computed on the image. In order to better follow salient structures of an image, one can replace the isotropic metric by an anisotropic metric that takes into account local orientations of image structures. A generalised formulation of anisotropic MPMs, can be defined as

$$E(\gamma) = \int_\gamma \Phi(\gamma'(s), T(\gamma(s)))|\gamma'(s)|ds \tag{2.14}$$

where $\gamma'(s)$ denotes the local orientation vector of the curve, and $T(\cdot)$ encodes the local anisotropy of the image. The function Φ is configured to encourage the tangent vector of γ to be collinear to the direction of local image structures. In practice, different anisotropic metrics of Φ have been defined depending on the applications, see examples in (Pichon et al., 2005, Peyré et al., 2010, Seong et al., 2008, Benmansour and Cohen, 2011). Generally speaking, the gradient of the image is taken into account, to measure the local directionality of the edge or texture.

2.2.3   Relation between ACM, MPM and Others

ACMs or snake models evolve an initially drawn curve in an image, to minimise an energy functional that usually consists of two terms: an internal

energy controlling the smoothness of the curve, and an external energy attracting the curve towards salient image features (e.g., edges). In models that directly stem from the original ACM, the energy functional takes into account both the length of the curve and its bending property, using first- and second-order derivatives with respect to s (see the definition in Equation (2.5)). This type of energy is however not intrinsic to the curve geometry, since it also depends on the parametrisation of the curve. Due to the parametrisation, a parametric active contour lacks the topology flexibility, and thus has the problem of simultaneous detecting/segmenting several objects. In (Caselles et al., 1997) and (Kichenassamy et al., 1995), the geodesic active contour (GAC), as a geometric version of the parametric ACM, was introduced that combines the internal and external energies of the snake model into a single term, which no longer depends upon the parameterisation of the evolving curve. Similar to snake models, GACs can have open and closed curves. Given that the minimisation of GAC based energies usually rely on the gradient descent strategy or an implicit level set based curve evolution, the evolving curve might be trapped in poor local minima of the energy, thus leading to a bad segmentation. For the case of a closed curve, to relieve the local minimisation problem, the extra pressure force (Cohen, 1991), GVF based forces and other related extensions have been introduced into the field. In the case of an open curve, Cohen and Kimmel (1997) introduced the minimal path approach that uses the Fast Marching propagation to find the global minimum of the associated energy. Their approach allows the user to specify two endpoints of a desired curve, and the same contour dependent energy as (Caselles et al., 1997) and (Kichenassamy et al., 1995) is constructed for the curve. Boundary constraints forbid the segmentation of a closed object with a single curve, but allow to track features such as roads in satellite imaging or vessels in medical imaging (Peyré et al., 2010). It is interesting that a strategy for detecting a closed curve has also been proposed in (Cohen and Kimmel, 1997). The approach allows to give only a starting point on the boundary of an object, and finds the complete closed contour as the union of two geodesic paths. Thus, the minimal path model is closely related to ACM and GAC models.

2.3   Conclusions

This chapter is concerned with the background algorithms and techniques that are related to the work presented in this thesis. An introduction to OF algorithms is provided, with details on how typical data terms and regularisation terms are constructed. Special attention is paid to the main research trends that are associated with the OF estimation, since improving OF based motion estimation is one of the major interests in this research. Specifically, in Chapter 3, a detailed literature review on OF regularisation approaches will be provided. Chapters 4 and 8 will propose two novel OF regularisation approaches respectively. In Chapters 6 and 7, the application of OF algorithms to cell tracking related tasks will be presented. After that, the standard formulations for the active contour model (ACM) and the minimal path model (MPM) have been described in this chapter. A short review on representative extensions of the ACM is provided, and, a comprehensive explanation is presented about how to design meaningful potential functions for MPMs, according to different application requirements. Additionally, the affinities between different types of energy-minimal contour models have been explained. The introduction of geodesic-distance based MPMs serves as preparation for Chapter 4. The ACM will be exploited as one of the main techniques in the proposed cell tracking framework in Chapter 7.

Chapter 3

Literature Survey and a Generalised Regularisation Formulation

Regularisation is a key tool to deal with the ill-posed problem[1]. Computationally, optic-flow (OF) estimation tries to find correspondences between pairs of pixels in either of two images. This requires matching pixels of similar intensity. Since merely imposing the matching of similar intensities will typically not give rise to a unique solution, OF estimation is thus a classical ill-posed problem (Horn and Schunck, 1981, Nagel and Enkelmann, 1986, Brox et al., 2004, Black and Anandan, 1996). To find an appropriate solution to this ill-posed problem, researchers have resorted to regularisation, which essentially introduces prior information about the underlying problem, hence making the search for feasible solutions.

The primary task of this chapter is to present an in-depth survey on existing OF regularisation approaches. With regard to this, regularisation approaches in the OF field are divided into two categories: the smoothness-prior and the sparsity-prior based approaches. Investigations range from classic approaches to recently developed approaches that represent the state of the art. Partic-

---

[1]A problem is said to be ill-posed in the Hadamard sense if it does not satisfy one of the following three conditions: existence, uniqueness and continuity (Hadamard, 1923)

ularly, much attention has been paid to non-local regularisation approaches, and higher-order derivatives based approaches (within the first category), and the sparsity-prior based regularisation strategy (corresponding to the second category), since they are closely related to the central contribution of this research. The relations between the two regularisation categories have also been established, for the first time, from a statistical point of view. After that, a thorough analysis and discussion is made about the common properties of the categorised approaches. This leads to a new, generalised formulation of OF regularisation schemes. Based on a further abstracting process, the potential research gaps are identified and several promising directions are suggested accordingly.

3.1 Standard Variational Optic-Flow Formulation

As stated in the last chapter, OF techniques that are formulated as variational, energy minimisation problems are referred to as global-energy type methods, where there is often an implicitly imposed, global-range smoothness constraint. OF methods that produce sparse flow fields (for some sampled feature points on an image) are usually referred to as local-energy type methods. In these methods, the motion vector of an individual feature is estimated independently of each other. The research of the current work mainly concentrates on regularisation approaches within the variational OF frameworks, which produce dense motion fields.

In order to compute a dense motion field $\mathbf{v} : \Omega \to \mathbb{R}^2$, it is common to define a variational energy functional that includes a data term and a regularisation term such as below:

$$E(\mathbf{v}) = \int_\Omega \Psi_{Data}(I(\mathbf{x} + \mathbf{v}) - I(\mathbf{x}))dxdy + \alpha \int_\Omega \Psi_{Prior}(\nabla \mathbf{v})dxdy. \quad (3.1)$$

where $\Psi_{Data}(\cdot)$ and $\Psi_{Prior}(\cdot)$ denote pre-selected penalisation functions. The most straightforward penalisation function is $(\cdot)^2$. ∇ stands for the gradient operation, thus $\nabla \mathbf{v} = (\mathbf{v}_x, \mathbf{v}_y)$. The data term aims to match pixels that obey pre-selected data-constancy assumptions, and the regularisation/prior term

(weighted by $\alpha > 0$) enforces prior knowledge about the underlying velocity field $\mathbf{v} = (u, v)^T$, such as motion coherence between a pair of nearby pixels in an image.

The state-of-the-art regularisation approaches in the OF field can be divided into two categories, namely the smoothness-regularisation based category, and the sparsity-prior based. The major difference between these two groups is that, the smoothness-assumption based group focuses on modelling motion coherence, and the other group aims to simultaneously model motion smoothness and discontinuities. The following sections present detailed investigations on the two categories of regularisation approaches.

3.2   Smoothness Regularisation

Smoothness regularisation plays an essential role in traditional OF frameworks, by enforcing spatial coherence properties on the pixel displacement field of an image. In detail, the spatial coherence constraint assumes that neighbouring pixels in an image have similar motion, i.e. are moving together and coherently. The assumption is based on a high-level understanding that neighbouring points in an image are likely to locate on the same object surface, where the motion of neighbouring points changes gradually.

The simplest smoothness regularity for estimating optic flows is defined inside the seminal work of Horn and Schunck (1981). The whole energy functional is defined as:

$$E(\mathbf{v}) = \int_{\Omega} (\nabla I^T \mathbf{v} + I_t)^2 dx dy + \int_{\Omega} \lambda (|\nabla u|^2 + |\nabla v|^2) dx dy. \tag{3.2}$$

where $\nabla I = (I_x, I_y)$ and I_t denote spatial and temporal partial derivatives of the image I, and T is used for the computation of matrix transpose. The symbol $| \cdot |$ denotes the total variation operator, and the positive parameter λ is to balance the first and the second terms in the energy equation. The spatial coherence constraint is formulated in terms of a least-squares estimation. This formulation assumes that the estimated flow corresponds to a single (continuously varying) motion field. Constrained by this quadratic-penalisation based formulation, the local flow vector of an image pixel is forced to be close to

the average of its neighbours. However, this assumption is incorrect at motion boundaries. Consider what happens if the flow field is discontinuous. There are multiple image motions in the neighbourhood. When a motion discontinuity is present, the quadratic regularisation term based coherence constraint will result in smoothing across the boundary, which reduces the accuracy of the flow field and obscures important structural information about the presence of an object boundary (Black and Anandan, 1996). This smoothness everywhere assumption is thus restrictive in some situations.

3.2.1   Robust Regularisation

To reduce the smoothing effect on motion boundaries, researchers have sought help from the statistical domain, in the sense that the spatial discontinuities in motion flows are treated as outliers in a statistical context. The traditional quadratic regularisation term has thus been recast in a robust estimation framework, which allows violations of the spatial coherence assumption that usually happen on motion boundaries. Here are some robust regularities with their corresponding penalisation functions:

- Truncated Quadratic (Black and Anandan, 1996):

$$
\rho(x, \alpha, \gamma) = \begin{cases} \gamma x^2 & \text{if } |x| < \frac{\sqrt{c}}{\sqrt{\gamma}}, \\ c & \text{otherwise.} \end{cases} \tag{3.3}
$$

  where $\gamma > 0$ controls the shape width of the x^2 function, and $c > 0$ determines the upper limit of the function.

- Geman&McClure (Black and Anandan, 1996):

$$
\rho(x, \sigma) = \frac{|\nabla x|^2}{\sigma + |\nabla x|^2} \tag{3.4}
$$

  where σ is the scale parameter.

- Lorentzian (Black and Anandan, 1996):

$$
\rho(x, \sigma) = log(1 + \frac{1}{2}(\frac{x}{\sigma})^2) \tag{3.5}
$$

  where σ is the scale parameter.

- Charbonnier (Bruhn et al., 2005):

$$\rho(x, \epsilon) = \sqrt{x^2 + \epsilon^2} \qquad\qquad (3.6)$$

where ϵ is a small positive value.

Combining robust-function based smoothness penalisation into OF estimation leads to the following objective functional:

$$E(\mathbf{v}) = E_{Data} + \alpha \int_\Omega \rho_S(\nabla\mathbf{v})dxdy \qquad\qquad (3.7)$$

with the same data fidelity term as in Equation (3.2)

$$E_{Data} = \int_\Omega (\nabla I^T \mathbf{v} + I_t)^2 dxdy \qquad\qquad (3.8)$$

and a robust smoothness term $\rho_S(\nabla\mathbf{v})$:

$$\rho_S(\nabla\mathbf{v}) = \frac{|\nabla u|^2}{\sigma + |\nabla u|^2} + \frac{|\nabla v|^2}{\sigma + |\nabla v|^2} \qquad\qquad (3.9)$$

The Geman&McClure function is taken for example here, while different formulations, as suggested by Black and Anandan (1996) and Sun, Roth and Black (2010), can also be chosen.

The problem with the least-squares solution is that outlying measurements are assigned a high "weight" by the quadratic smoothness term. While, the robust regularisers are designed to be more forgiving about outlying measurements. The most simple robust regularity is the truncated quadratic. As illustrated in Figure 3.1, up to a fixed threshold, local motion un-smoothness is penalised quadratically, but beyond that, a constant value of penalisation is assigned.

Among the listed robust regularities, the Charbonnier based function (with a small parameter $\epsilon > 0$) is a special one, since it approximates the total-variation (also called L1-norm) based penalisation (Rudin et al., 1992b): $TV(u) = \int_\Omega |\nabla u| dx$, which is very popular for regularising inverse problems (Zach et al., 2007). Although the total-variation (TV) based penalisation function appears to be relatively simple, it offers some computational difficulties. This is mainly due to the fact that it is not continuously differentiable. In practice, some researchers choose a slightly compromised way of dealing with that fact by approximating it with the Charbonnier function (Brox et al., 2004). Others
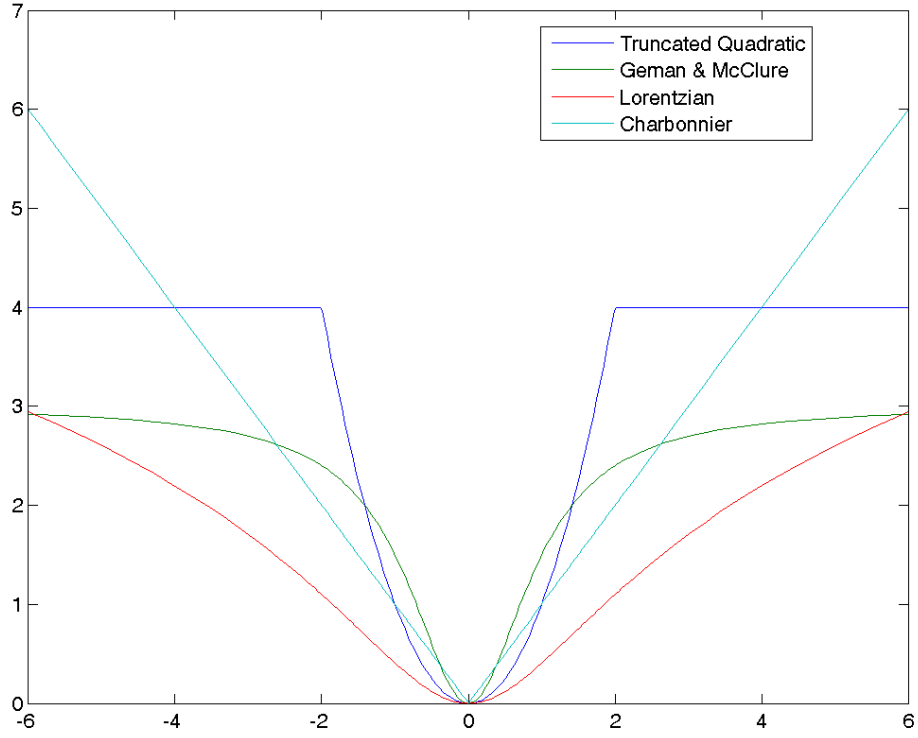
Fig. 3.1: Different robust functions for spatial smoothness terms: Truncated Quadratic ($c = 4, \gamma = 1$), German & McClure ($\sigma = 1$), Lorentzian ($\sigma = 1$), and Charbonnier ($\epsilon = 0.001$).

employ an alternative strategy, such as that proposed by Chambolle (2004), to solve the Rudin-Osher-Fatemi energy (Rudin et al., 1992b) for TV involved energy minimisation.

Although different robust-function based regularities exist in the literature, their common property is the ability of reducing the effect of outliers.

3.2.2 Anisotropic Regularisation

In order to facilitate the distinguishing of isotropic and anisotropic regularities, the Horn-Schunck OF functional (Equation (3.2)) has been re-written as below

with a discrete smoothness term:

$$E(u,v) = E_{Data} + \sum_{\mathbf{x} \in \Omega} \sum_{\tilde{\mathbf{x}} \in N_{\mathbf{x}}} \alpha(|u_{\mathbf{x}} - u_{\tilde{\mathbf{x}}}|^2 + |v_{\mathbf{x}} - v_{\tilde{\mathbf{x}}}|^2) \tag{3.10}$$

where E_{Data} denotes a standard data conservation term (as presented in Equation (3.8)). $N_{\mathbf{x}}$ represents the set of the four or eight directional neighbours of pixel $\mathbf{x}$. Notice the $\alpha > 0$ is a constant parameter weighting the spatial coherence between pixel $\mathbf{x}$ and its neighbour $\tilde{\mathbf{x}}$ (with $\tilde{\mathbf{x}} \in N_{\mathbf{x}}$). The constantly weighted regularity allows the pixel $\mathbf{x}$ to get motion support evenly from its local neighbours in all orientations, and is thus referred to as isotropic regularity. This type of regularisation is reasonable when the neighbours within the set $N_{\mathbf{x}}$ all locate on the same motion surface with $\mathbf{x}$. It is not suitable when $N_{\mathbf{x}}$ contains pixels from two or more separate motion regions, where motion boundaries exist.

In order to respect motion boundaries, the so-called anisotropic regularities have been proposed for taking image discontinuities into account. Alvarez (1999) proposed a model with a scalar-valued weight function that reduces the regularisation at image edges. An anisotropic counterpart that exploits the directional information of image discontinuities was introduced by Nagel and Enkelmann (1986). Their method regularises the flow along image edges but not across them, which basically assigns more motion constraint between pixels with similar image gradients. This type of motion coherent constraint is also called "oriented smoothness". Xiao et al. (2006) proposed n adaptive smoothness weight that is calculated by the Bilateral filter scheme, where the motion coherence degree between pixels $\mathbf{x}$ and $\tilde{\mathbf{x}}$ is decided by not only the spatial proximity but also their brightness or colour similarity. One interesting point within (Xiao et al., 2006) is that, an image pixel is permitted to connect with not only its directly adjacent neighbours, but also some of its non-adjacent pixels within an slightly enlarged neighbourhood. By doing so, mutual motion supports between non-adjacent pixels are established. The work of Xiao et al. (2006) thus plays a meaningful role in fostering the development of the non-local scale of OF regularisation approaches. Zimmer et al. (2011) perform an anisotropic weighting scheme that is similar to the "oriented smoothness", but the directions are defined by the data constraint rather than the image gradient.

Since the smoothness weights in the aforementioned isotropic regularities are adaptively determined, according to the measurements of pixel colour or gradient similarities, these regularities are also called image-driven regularisation strategies. Generally speaking, these image-driven regularities assume that discontinuities of the motion field tend to coincide with object boundaries and discontinuities of the brightness function.

Of course, not every image edge will coincide with a flow edge. Thus, the image-driven strategies, are prone to resulting in over-segmentation artefacts in textured image regions. To avoid this, flow-driven regularisers have been proposed that respect discontinuities of the evolving flow field, and are therefore not misled by image textures (Zimmer et al., 2011). The drawback of flow-driven regularisers lies in less well-localised flow edges, compared to image-driven approaches. Concerning the individual problems of image- and flow-driven strategies, ideas arise to combine the advantages of both types of approaches (Zimmer et al., 2011).

The equation below presents a general anisotropic-regularity based smoothness term,

$$E_{Smooth}(u,v) = \sum_{\mathbf{x}\in\Omega} \sum_{\tilde{\mathbf{x}}\in N_{\mathbf{x}}} \alpha_{\mathbf{x},\tilde{\mathbf{x}}}(|u_{\mathbf{x}} - u_{\tilde{\mathbf{x}}}|^2 + |v_{\mathbf{x}} - v_{\tilde{\mathbf{x}}}|^2), \tag{3.11}$$

where the weighting value α might be determined according to some pairwise affinity measurement, for example:

$$\alpha_{\mathbf{x},\tilde{\mathbf{x}}} = \exp\{-\frac{|\mathbf{x} - \tilde{\mathbf{x}}|^2}{2\sigma_1} - \frac{|I(\mathbf{x}) - I(\tilde{\mathbf{x}})|^2}{2\sigma_2} - \frac{|\mathbf{v}(\mathbf{x}) - \mathbf{v}(\tilde{\mathbf{x}})|^2}{2\sigma_3}\} \tag{3.12}$$

where σ_1, σ_2 and σ_3 are respectively the distance variance, colour variance, and flow variance of adjacent pixels around $\mathbf{x}$. By this means, the regularisation weights are adaptively changed, according to some prior knowledge, depending on which neighbouring pixels are more likely located on the same surface with the pixel $\mathbf{x}$, and which pixels thus deserve higher weights. Existing examples of prior knowledge for choosing suitable motion supporters/integrators[2] for a pixel include pairwise similarities of spatial location, colour, gradient, motion, etc.

---

[2]In this work, a pixel's motion supporters/integrators are a group of local/non-local neighbouring pixels, whose OF vectors the motion of the considered pixel is constrained to be similar to.

3.2.3   Non-Local Motion Regularisation

Considering OF estimation requires the pooling of constraints over some spatial neighbourhood, mutual motion constraints are imposed within a relatively local image region in traditional regularisation approaches. Recent works, however, have advocated the necessity of introducing larger/non-local range of motion regularisation. This is due to the experimental observation that traditional local region regularisation is usually not sufficient to constrain the solution (Krähenbühl and Koltun, 2012, Sun, Roth and Black, 2010, Werlberger et al., 2010). Figure 3.2 illustrates local and non-local ranges of motion coherence relationships for the same pixel.
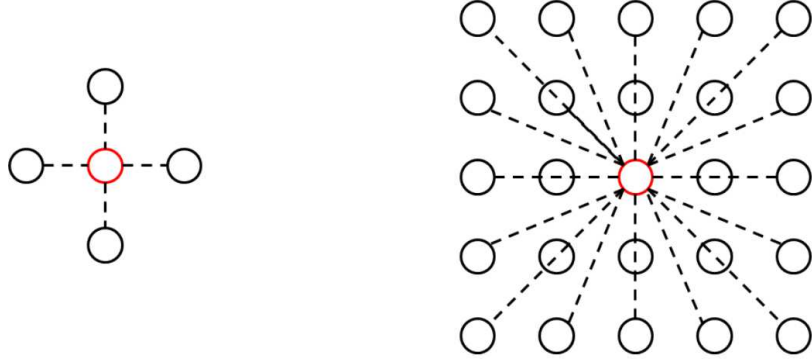
Fig. 3.2: Neighbours for the pairwise model (left) and the non-local model (right). The pairwise model connects a pixel with its nearest neighbours, while the non-local term connects a pixel with many pixels in a large spatial neighbourhood (Sun, Roth and Black, 2010).

As mentioned in the last section, within (Xiao et al., 2006), a slightly enlarged range of motion integration region is enabled. The strategy is mainly designed for pixels locating in motion occlusion regions, so the occlusion-region pixels can expect to borrow motion information from non-occluded regions on the same surface. Their work has a focus on recovering an occlusion region's motion, but actually has not introduced a non-local motion constraint for every pixel inside the image. Therefore, The work of Xiao et al. (2006) can be considered as a transition between local and non-local ranges of OF regularisation approaches.

The non-local smoothing strategy was earlier proposed by Yaroslavsky and Yaroslavskij (1985) and has recently been introduced to the OF community by Sun, Roth and Black (2010) and Werlberger et al. (2010). In these approaches, the motion information can be shared between pixel pairs within a large spatial neighbourhood. Theoretically, a fully-connected graph that establishes mutual motion supporters for any pixel pairs within the image is allowable (Krähenbühl and Koltun, 2012). Therefore, the motion estimation accuracy can be largely improved within particularly difficult situations, including large textureless regions where no information is available, and occlusion regions where the data conservation assumption is violated.

The equation below demonstrates how a non-local smoothness term is usually formulated:

$$E_{Smooth}(u,v) = \sum_{\mathbf{x} \in \Omega} \sum_{\tilde{\mathbf{x}} \in N_{\mathbf{x}}^{NL}} \alpha_{\mathbf{x},\tilde{\mathbf{x}}}(|u_{\mathbf{x}} - u_{\tilde{\mathbf{x}}}|^2 + |v_{\mathbf{x}} - v_{\tilde{\mathbf{x}}}|^2), \qquad (3.13)$$

where, $N_{\mathbf{x}}^{NL}$ denotes the non-local, motion supporter region for pixel $\mathbf{x}$. This equation is very similar to the local smoothness term as defined in Equation (3.11), with only a different scale of motion supporter range. In principle, since the non-local range can be a pre-selected large region, or even the whole image domain, adaptive weighting is necessary to guide how much motion correlation should be established between any involved pair of pixels. Similarly to how an adaptive smoothness weight is determined within image- or flow-driven regularisation schemes, the value of $\alpha_{\mathbf{x},\tilde{\mathbf{x}}}$ can be defined as in Equation (3.12).

In existing non-local based OF approaches, the Bilateral filter based strategy is a popular choice. Another techniques, such as the Intervening Contour based pairwise affinity measurer (Ren, 2008), have also been suggested. More details for comparing these strategies will be delivered in Chapter 4. Essentially, how to select a suitable set of non-local motion supporters for a pixel involves a non-local, low-level image segmentation process. In that sense, pixels with similar colour, that are more likely to locate on the same surface, should be grouped together to mutually support each other.

One challenging problem of non-local regularisation based approaches is that, the computational complexity grows quadratically with the maximal distance of the non-local connection that is allowed for each pixel. In order to deal with

this issue, Krähenbühl and Koltun (2012) proposed an efficient framework that approximates the non-local regularities with high-dimensional Gaussian filters, to optimise the energy minimisation process.

In spite of the improvements that non-local regularisation based approaches have brought into the OF field, one intrinsic problem exists. That is, the larger the region of motion integration is allowed, the more likely it is to contain non-constant motion, such as multiple motions from different objects, and non-rigid motion that violates the constant or the affine motion constraints.

3.2.4   Higher-Order Derivatives based Regularisation

One common property of the aforementioned local and non-local smoothness terms is that they are all based on the penalisation of first-order motion derivatives or, more generally speaking, pairwise motion deviations. However, pairwise-similarity based motion coherence constraints restrict the motion model to translocation motion. Other types of motion patterns, such as rotation or scaling, will be penalised(Kwon et al., 2010), not to mention non-rigid motion based transformation.

To deal with this problem, incorporating higher-order priors into OF frameworks has recently come into the spotlight. For example, replacing the first-order smoothness priors with the ones that encourage the second-order derivatives (such as $(\frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial y^2}, \frac{\partial^2 u}{\partial x \partial y})$), or even higher-order derivatives to be small (Baker et al., 2011). By doing so, physically sound properties of the flow can be exploited as prior smoothness constraints, which encourage the estimated motion to follow the pre-assumed flow pattern.

Lee et al. (2010) proposed a regularisation term that penalises the deviation between original flow and a modified Bilateral filtered flow. By doing so, the traditional, pair-wise constrained motion coherence is replaced by a group-wise constrained one. Their approach thus implicitly involves a higher-order smoothness prior.

To reduce the motion estimation bias towards piecewise constant motions, Trobin et al. (2008) proposed a second-order derivative based regularisation ap-

31

proach, replacing the original first-order based one. The approach has demonstrated improved performance in recovering motions for weakly textured areas.

Kwon et al. (2010) proposed Markov Random Field based energies with second- and third-order smoothness priors. They have proved that higher-order smoothness based priors can lead to better results than first-order based approaches in various applications, including stereo matching and non-rigid image registration (both problems are strongly related to OF estimation).

Yuan et al. (2007) designed a method that simultaneously estimates and decomposes OF fields for non-rigid motions. They suggest that higher-order regularisation is able to incorporate physical properties, with respect to the characteristics of a particular type of motion. So, this type of OF methods can accurately recover meaningful flow structures.

In the work of Onkarappa and Sappa (2013), a Laplacian derivative (involving second-order motion derivatives) based regularisation is proposed, that replaces traditional, motion gradient (first-order derivative) based smoothness terms in OF estimation. The proposed work has achieved improved performance in recovering motion fields for driving scenarios that have large variations in speed and rotations (Onkarappa and Sappa, 2013).

Li et al. (2013) proposed a Laplacian Mesh Deformation (LMD) based smoothness prior for OF estimation. The introduced constraint is imposed on the Laplacian coordinates, and can encourage local regularity of the involved mesh whilst allowing global non-rigidity. If the motion field is considered as a 3D surface profile, the LMD based constraint penalises the mean curvature of the motion "surface". Thus, this constraint is intrinsically a second-order derivatives based regularity.

Ranftl et al. (2012) proposed a signal regulariser based on the second-order case of Total Generalised Variation (TGV-2) (Bredies et al., 2010), for the purpose of stereo matching. The TGV-2 based smoothness term brings the discontinuity preserving quality of the total-variation based regularity to second-order derivatives. So, instead of favoring a motion field with piecewise constant solutions, TGV-2 regularities fit well with piecewise affine solutions.

Kadri-Harouna et al. (2013) designed a higher-order regularisation scheme that transforms high-order derivative calculations as the process of penalising the amplitude of small-scale wavelet coefficients. The rationale behind their approach is that, for a given motion signal, the behaviour of its small-scale coefficients that result from an n vanishing-moment wavelet decomposition can be related to the nth derivative of the signal; so, the penalisation of small-scale coefficients' amplitude enables the control of the derivatives of the estimated motion (Kadri-Harouna et al., 2013). By this means, the difficult problem of directly calculating high-order derivatives is circumvented, and a relatively efficient high-order regulariser that is associated to a low computational complexity is accomplished. Moreover, the authors also introduced a divergence-free constraint, as a smoothness prior into their wavelet based approach, which has demonstrated its advantage in estimating turbulent fluid motions.

In addition, researchers suggest the use of more sophisticated regularisation techniques that are based on statistical learning, which adopt learned, high-order motion priors to guide OF estimation. Roth and Black (2007) learned the spatial statistics of optic flow, which was shown to be heavy-tailed. Then, they used the learned prior model to regularise flow estimation. In their work, spatial iterations are considered within a 3×3 pixels' scale. Sun et al. (2008) learned statistical models of image structure-adaptive flow derivatives, and exploited the models in motion regularisation.

3.2.5   Layer, Segmentation based Regularisation

Different from previous regularities that exploit non-parametric motion models, layer or segmentation based approaches usually assume a parametric motion model, e.g. translocation or affine model, for each segment. Taking the affine-model based regularisation for example, an affine motion model can be defined and parameterised by $\mathbf{a}_s = (a_{s0}, a_{s1}, a_{s2}, a_{s3}, a_{s4}, a_{s5})^T$, which represents the vector of all affine parameters for a segment s (Xu et al., 2008). The motion field $\mathbf{w}(\mathbf{a}_s, \mathbf{x})^T$ in segment s is thus given by

$$u(\mathbf{a}_s, \mathbf{x}) = a_{s0}x + a_{s1}y + a_{s2}, \tag{3.14}$$

$$v(\mathbf{a}_s, \mathbf{x}) = a_{s3}x + a_{s4}y + a_{s5}. \tag{3.15}$$

A typical way of recovering the affine parameters is a regression based approach, which minimises the following objective function:

$$E(\mathbf{a}_s) = \sum_{\mathbf{x} \in s} ((\nabla I)^T \mathbf{w}(\mathbf{a}_s, \mathbf{x}) + I_t)^2. \tag{3.16}$$

According to the recovered affine vector, the motion flow within the segment s will be determined.

In fact, parametric and non-parametric models based regularisation approaches are linked to each other (Myronenko, 2010). Although discriminative formulations are exploited, and the number of parameters to be estimated are different, the two types of approaches both intrinsically impose motion-coherence constraints.

Common segmentation based approaches exploit super-pixels (Xu et al., 2008, Zitnick et al., 2004), that are usually obtained from image over-segmentation, as image segments. Originally, in order to avoid including multiple motions or non-rigid motions that violate the affine motion assumption, the size of the involved patches or super-pixels has been chosen to be small (Xu et al., 2008).

However, the utilisation of a small area of motion integration may result in poor estimation of the motion parameters, because of the local aperture problem. In order to deal with the issue, Xu et al. (2008) add an inter-segment smoothness constraint that can spread motion information between adjacent super-pixels. By doing so, the local affine motion constraint is enforced, and simultaneously a globally relaxed deformable field is also enabled. An alternative direction to relieve the local aperture problem is expanding the segments' area, which is in spirit very similar to the non-local range regularisation. In relation to this, layer based regularisation has been proposed, where the layered model is like a cardboard cutout representation of an image. Layered models have many advantages: explicitly modelling different object surfaces, that may overlap and occlude each other, so making the detection of occlusion boundaries possible; and offers an elegant approach to motion segmentation. Whereas, by allowing considerably large area of segments, a single parametric motion model for each layer will be too restrictive to capture the motion of natural scenes. In order to relax the constraint, Black and Jepson (1996) add local deformation to the parametric model. Mémin and Pérez (2002) combine

the piece-wise parametric motion with local disturbance in a hierarchical setting, to mix the local flow field with different parameterisations. Weiss (1997) addresses this by allowing smooth motions in the layer. Sun, Sudderth and Black (2010), Sun et al. (2012) impose a semi-parametric model as a motion coherence constraint. Their model encourages the flow field of each layer to be close to a pre-estimated affine flow, whilst allowing deviations from the affine motion, by using a robust function (as introduced in Section 3.2.1), to penalise the difference between the estimated flow and the predicted affine flow.

3.2.6 Other Smoothness Regularisation Approaches

Not all regularisation approaches applied in OF algorithms can be described in terms of the previously listed categories and representations. Here some additional approaches and representations are briefly mentioned, considering that they will be later used in this research work.

Contour based Regularisation: Liu et al. (2006) proposed a contour based, motion estimation approach, in which only object boundary features are taken into account for data matching, and the motion coherence constraint is enforced along contours. The advantage of this regularisation scheme is manifested in estimating motions for objects with no visible texture. In this situation, it is often difficult to determine the reliability of motion estimation from local measurements. Properly grouping boundary fragments into contours and enforcing motion smoothness along contours has proved the advantage of disambiguating the uncertainties of local estimates.

Global Regularisation Imposed on Sparsely Sampled Features

Most variational OF algorithms are applied on densely gridded pixels inside an image, such as those introduced in the previous sections. There are also a few of the variational OF frameworks that take into account only the sparsely sampled features (usually texture and/or edge features). Similarly to how a global smoothness constraint is implicitly imposed in the former type of OF algorithms, the motion-vector estimation for each sparsely sampled feature is directly/indirectly constrained by other sampled features within the image. The

regularisation embedded within the latter type of OF algorithms is referred to as sparsely-sampled features based global regularisation. Correspondingly, the regularisation strategies associated with the former type of OF algorithms can be referred to as densely-gridded pixels based global regularisation approaches.

Although these two types of regularisation both stem from the regularisation scheme of Horn and Schunck (1981), the sparsely-sampled features based strategy in fact enables a non-local range of explicit motion regularisation. The concept is in spirit similar to the non-local scale of OF regularisation as introduced in Section 3.2.3. The advantage of the sparsely-sampled features based regularisation is that, it is more flexible in selecting motion supporters for a pixel/feature along meaningful image structures. For example, Liu et al. (2006) proposed a contour based regularisation approach, in which the motion supporters for a boundary feature is selected along the boundary contour. Another example is proposed by Birchfield and Pundlik (2008), where only texture and edge features are taken into account, and the group-wise motion coherence is imposed on the sparsely sampled features.

Temporal Regularisation

The OF approaches as discussed so far are designed to estimate two-frame motion flows. The associated smoothness terms model spatially smooth flow fields. When considering an image sequence or a video that in general encompasses more than two frames, it makes sense to also assume a temporal smoothness of the flow sequence. This leads to temporal regularisation strategies (Zimmer et al., 2011), which usually assume $(\frac{\partial u}{\partial t})$ and $(\frac{\partial v}{\partial t})$ are small, with t denoting the temporal-axis associated variable. More examples for temporal regularisation approaches can be found in (Murray and Buxton, 1987, Nagel, 1990, Weickert and Schnörr, 2001, Zimmer et al., 2011, Brox et al., 2006).

3.3   Sparsity Priors based Regularisation

An alternative approach to incorporate prior knowledge about motion signals is via sparse representation. Sparsity based motion estimation assumes that the OF field is sparse in some domain. The domain is spanned by a set of bases

or a dictionary of atoms (Jia et al., 2011). So, the motion signal $\mathbf{v}$ can be well approximated by a small set of selected atoms from the dictionary. This assumption is reasonable because, although the flow patterns may be complex and varying across the whole field, they are much simpler compared with the signal patterns of natural images[3] (Shen and Wu, 2010). From the perspective of compressive sensing[4], sparsity priors based OF estimation amounts to recover a dense flow field from much fewer measurements, thus solving the aperture problem (Jia et al., 2011). Unlike the traditional smoothness constraint, the proposed sparsity constraint handles motion homogeneities and motion discontinuities at the same time.

3.3.1 Different Basis based Sparsity Priors

Recently, researchers are starting to exploit sparse representations of the motion field for regularising OF estimation, and have already achieved encouraging results in the works of Shen and Wu (2010), Nawaz et al. (2011), Han et al. (2011), Jia et al. (2011), Chen et al. (2012), Ottaviano and Kohli (2013). The introduced sparsity priors assume that the flow field can be sparsely represented in certain spaces/domains, such as gradient field domains, wavelet and DCT domains.

Shen and Wu (2010) assume that the flow field is sparse in both the Haar wavelet domain and the gradient domain, i.e. a large number of the Haar wavelet coefficients as well as the gradients of the flow field are very small. Nawaz et al. (2011) and Han et al. (2011) also take advantages of the sparsity of motion gradient fields. In the work of Ottaviano and Kohli (2013), a logarithmic-penalty based regularity is applied on the wavelet transformed coefficients of the motion field, to encourage the motion field to be sparse and thus compressible. Chen et al. (2012) exploit the gradient-domain based sparsity prior. An interesting point is that the authors decompose the motion field

---

[3]Here, a natural image means a digital photograph of a real-world scene.

[4]Compressed sensing (also known as compressive sensing) is a signal processing technique for efficiently acquiring and reconstructing a signal, by finding solutions to under-determined linear systems. This takes advantages of the signal's sparseness or compressibility in some domain, allowing the entire signal to be determined from relatively few measurements.

into sparse and non-sparse components. In more detail, they impose sparseness on the motion-discontinuity component of the flow, and a density prior on the residual component. In addition, learning-based sparsity priors have also been considered for regularising motion estimation. For example, Jia et al. (2011) train an over-complete OF dictionary that adapts to the ground flow fields, and assumes that each flow patch can be encoded via a sparse representation over the trained flow dictionary.

3.3.2 Sparsity Prior, L1-Norm, and Laplacian Distribution

As previously mentioned, sparsity is a property which can be described either directly on the original motion field or in a transformed space, for example on the derivative of the motion field, or more generally, on the coefficients of the projection of the motion signal on any basis or any set of functions.

It is common to use the L0-norm $L_0(f) = \|f\|_0$, or the L1-norm $L_1(f) = \|f\|_1$, to enforce the sparsity of the solution (Donoho, 2006, Candès et al., 2008). The generalisation of L_p norms is written as,

$$L_p(f) = \Big(\sum_j |f_j|^p \Big)^{\frac{1}{p}}, p \geq 0 \tag{3.17}$$

Since the L0-norm can be difficult to optimise, due to its discrete and combinatorial nature (He and Schaefer, 2013), there is a large volume of works that choose to relax the L0-norm with an L1-norm in their implementations, due to the consideration that the L1-norm gives the closest convex optimisation to the L0-norm, and can be solved by representative L1-minimisation strategies, such as listed in the work of Yang et al. (2013). For this reason, the L1-norm has become the established mechanism with which to promote sparsity in many applications. In the past two decades, there has been significant interest in the L1-norm based regularisation methods, including sparse wavelet representations and the total-variation regularisation. This is because of the appealing idea of creating accurate predictive models that also have interpretable or parsimonious representations (Mohamed et al., 2012).

The L1 norm and the Laplacian distribution (in the statistical domain) are

closely related. Consider a L1-norm based regularisation term, defined as

$$E_{L1}(f) = \lambda \sum_j |f_j|, \quad f = \operatorname*{argmin}_f E_{L1}(f) \tag{3.18}$$

where λ is an associated weighting parameter for the L1-norm based regularity. Notice that $|f_j|$ is proportional to the minus log-density of the double-exponential distribution, i.e. Laplacian distribution. As a result, $|f_j|$ can be interpreted as the following Bayesian posterior model (Tibshirani, 1996),

$$P(f_j) = \frac{1}{2\tau} exp\left(- \frac{|f_j|}{\tau} \right) \tag{3.19}$$

with $\tau = \frac{1}{\lambda}$. So, the L1-norm can be derived using the posterior mode analysis, when the parameters f_js have independent and identical Laplacian priors,

$$E_{L1}(f) \propto -log\left(\frac{1}{2\tau} exp\left(- \frac{\sum_j |f_j|}{\tau} \right) \right) \tag{3.20}$$

Therefore, if the latent distribution of f is Laplacian, the maximum a posteriori solution for f is equivalent to the L1-norm regularisation in this model.

3.4   Summaries

Table 3.1: The categorisation of representative OF regularisation approaches

Regularisation approaches	References	MSR	MD	SoPMI
Robust-func based	(Black and Anandan, 1996, Zach et al., 2007, Brox et al., 2004, Sun, Roth and Black, 2010, Werlberger et al., 2010)			X
Anisotropic smoothness	(Nagel and Enkelmann, 1986, Alvarez, 1999, Xiao et al., 2006, Sun, Roth and Black, 2010, Werlberger et al., 2010, Zimmer et al., 2011, Liu et al., 2006)	X		
Non-local	(Sun, Roth and Black, 2010, Werlberger et al., 2010, Krähenbühl and Koltun, 2012)	X	X	
Segmentation based	(Black and Jepson, 1996, Weiss, 1997, Mémin and Pérez, 2002, Zitnick et al., 2004, Xu et al., 2008, Sun, Sudderth and Black, 2010, Sun et al., 2012, 2013, Liu et al., 2006)	X	X	
f.o.d smoothness	(Black and Anandan, 1996, Nagel and Enkelmann, 1986, Alvarez, 1999, Xiao et al., 2006, Liu et al., 2006, Ren, 2008, Zach et al., 2007, Brox et al., 2004, Sun, Roth and Black, 2010, Werlberger et al., 2010)		X	X
Higher-order smoothness	(Kwon et al., 2010, Lee et al., 2010, Yuan et al., 2007, Onkarappa and Sappa, 2013, Li et al., 2013, Ranftl et al., 2012, Kadri-Harouna et al., 2013, Roth and Black, 2007, Sun et al., 2008, Jia et al., 2011, Trobin et al., 2008, Xu et al., 2008)	X	X	X
Gradient field sparsity	(Shen and Wu, 2010, Nawaz et al., 2011, Han et al., 2011, Chen et al., 2012)		X	X
Wavelet domain sparsity	(Shen and Wu, 2010, Ottaviano and Kohli, 2013, Jia et al., 2011)		X	X

Table 3.1 presents a summary of most of the previously introduced regularisation approaches. The table also categorises the associated references into different groups, according to the considered criteria, namely, whether the factors of motion supporter regions, motion spaces, and motion-distribution priors are considered. On one hand, the groups of the methods have distinctive characteristics, in terms of the criteria. On the other hand, it is easy to understand that some groups of regularisation approaches contain overlapping properties. The essential similarities between the regularisation groups have motivated a further analysis. The three issues, namely selecting a motion supporter region (MSR)[5], choosing motion domains or spaces, and choosing motion-distribution priors, that are of crucial importance to specifically determine a regularisation scheme, have been revealed as follows:

1. Selecting a **motion supporter region (MSR)**. Many regularisation strategies need to solve the problem of selecting suitable motion supporters for each pixel/feature in motion estimation. Table 3.1 marks the regularisation groups where selecting suitable MSRs is heavily involved (with the symbol X), including: the anisotropic group, the non-local group, the segmentation based, and the higher-order smoothness based regularisation groups. A common property is that they focus on the modelling of motion smoothness, and exclude situations of motion discontinuities.

2. Choosing **motion domains (MDs) or spaces** to pose regularisation constraints on. Within existing OF regularisation approaches, the following groups of motion domains exist:

- Motion derivatives based domains: 1) The domains obtained from derivative operators that are conducted on locally connected maps (LCMs)[6] of the flow field. These include the gradient field of the flow (i.e., the most

---

[5]In this work, a motion supporter region (MSR), or alternatively called motion integration region (MIR), is a local/non-local neighbouring region surrounding a considered pixel; during the optic-flow estimation, the motion vector of the considered pixel is constrained to be similar to those of the pixels in this neighbouring region, which is referred to as the pixel's motion supporter region.

[6]In this research work, the locally connected map (LCM) represents the graph that consists of the vertices of the gridded pixels of a field, and edges that are determined by connecting immediately adjacent vertices.

common motion space), second-order derivatives (e.g., the Hessian and the Laplacian) and high-order derivatives induced fields of the flow; 2) Domains of motion derivatives that are calculated on the non-locally or densely connected map (DCM)[7], e.g., within non-local regularisation approaches (see (Krähenbühl and Koltun, 2012, Sun, Roth and Black, 2010, Werlberger et al., 2010)), the constraint of this type of motion coherence is enforced on non-local, first-order derivatives of the flow field.

- Motion domains or spaces that are resulting from wavelet transformations, such as the domains that are induced by the Haar function, or other B-spline functions based wavelet transformations.

- Parametrisation-models (PMs) induced spaces: PMs such as affine, rigid, translocation, and constant motion models have been exploited for OF estimation, see examples in (Xu et al., 2008, Nir et al., 2008). By considering PMs induced motion domains, instead of directly penalising changes in the optic flow or the wavelet coefficients of the flow field, PMs based regularisation approaches accumulate a cost of deviating from the assumed, parameterised motion model. Notice that regularisation approaches based on PMs are closely related to derivatives based regularisation methods, so, the focus of this research is mainly on derivatives and WTs based motion domains. Graph 3.1 gives a summary of the aforementioned motion domains and their relationships.
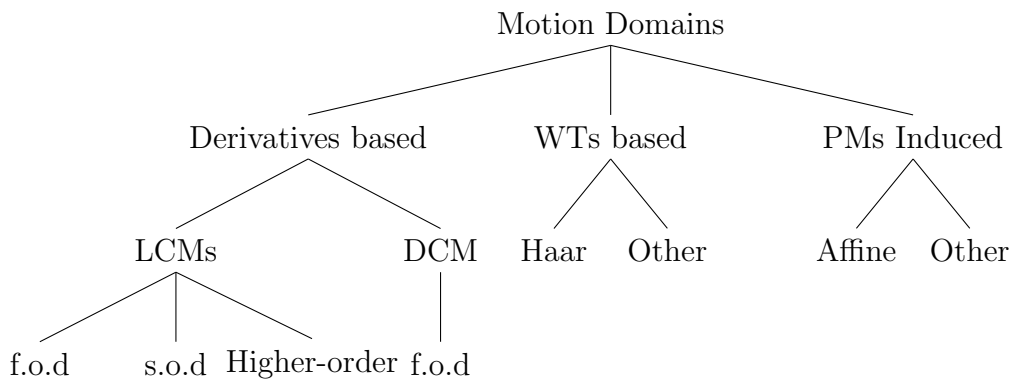
Graph 3.1: Categorisation of Motion Spaces.

---

[7]In this thesis, the densely connected map (DCM) represents the graph that consists of the vertices of the gridded pixels of a field, and edges that are established by connecting any of two vertices in the graph.

3. Making assumptions about the **statistics of the projected motion information (SoPMI)**. From the first variational OF algorithm that exploits the Tikhonov based regularisation, to the well-known, robust-function based smoothness penalisers, and the very popular TV (or L1-norm) based regularisation approaches, statistical distribution assumptions are explicitly or implicitly made about the projected motion information (that is obtained after some space transformations from the original motion field). For example, when used together with the first-order derivative (f.o.d) induced motion spaces, the Tikhonov and the TV based penalisation functions will encourage the motion-gradient field to respectively obey the Gaussian distribution, and the Laplacian distribution. These two distributions both belong to the exponential distribution family, and so, they can be uniformly categorised into the generalised Gaussian distribution. For other penalisation functions, the corresponding statistical distributions have been mentioned in the review paper of Mohammad-Djafari (2012).

It can be understood from the existing literature that all regularisation approaches in the variational OF frameworks explicitly or implicitly deal with the aforementioned three aspects. The second and the third aspects are usually jointly considered as a unified problem in the design of motion prior models. Compared with the integral understanding of a motion regularisation scheme, a decomposed interpretation that considers a motion regularity to consist of three sub-components, namely MSRs, motion spaces, and motion statistics of the motion space, has an important merit. This is believed to further reveal the underlying logic of the regularisation theory, and hence promote more improvements in the regularisation field.

3.5 Understanding the Connections between Different Regularities

3.5.1 Relationships between Motion Spaces

Derivative operators and wavelet transformations (WTs) based functions/basis are commonly used to induce desired motion spaces, in non-parametric regularisation approaches, for OF estimation. Due to the fact that derivatives based regularisation techniques put an emphasis in modelling motion smoothness (treating discontinuities as outliers), and sparsity priors based regularities are claimed to simultaneously model homogeneous and discontinuous motion flows, one may consider that derivatives based and WTs based motion spaces are different. The close relationship between the two groups of signal (image and motion) spaces has been studied by Steidl and Weickert (2002) and Cai et al. (2012), and can be also reflected in the works of Kadri-Harouna et al. (2013), Shen and Wu (2010), Nawaz et al. (2011) and Han et al. (2011).

Steidl and Weickert (2002) studied the mathematical equivalence between the total variation (TV) and the WT under some special condition. Cai et al. (2012) remarked that the WT (by choosing appropriate basis) can be used to approximate different orders of (local) differential operators. In the work of Kadri-Harouna et al. (2013), for a given motion signal, the coefficients that result from an n vanishing-moment, wavelet decomposition is related to the nth (local) derivative of the signal. Nawaz et al. (2011) and Han et al. (2011) have applied the f.o.d based motion space (a standard space in smoothness-assumption based regularisation approaches) into the sparsity prior based regularisation schemes. A special point of Nawaz et al. (2011), Han et al. (2011) is directly imposing sparsity priors on the motion-gradient field, rather than using the more sophisticated but standard WTs based motion spaces. From the sparsity based regularisation view, these works transform the original motion field using a f.o.d basis function. From the smoothness based regularisation angle, the works can be considered as combining the f.o.d based motion space with the TV (or L1-norm) based penalisation function. This interpretation gives an interesting clue that, when collaborating with the TV based penali-

sation function, a derivative based motion space may lead to a sparsity prior based regularisation scheme.

Because of the aforementioned works, the close connection between the derivative based and the WT based motion spaces has been revealed: the derivative based motion spaces can be usually approximated by WT induced spaces, and WTs can potentially produce a broader range of signal spaces than differential operations.

To sum up, it can be understood that choosing motion bases in OF regularities corresponds to exploring different types and ranges of repetitive information, or more essentially, exploiting the signal compressibly within the considered motion space. Specifically, the smoothness assumption based regularities utilise the motion-information redundance in the original OF field or motion spaces that are induced by low-/high-order differential operations. And, the sparsity assumption based approaches also utilise the motion redundance in the considered WT induced motion spaces.

3.5.2 Relationships between Motion-Distribution Priors

OF regularisation approaches usually require a pre-determined motion space, on which further constraints can be applied. Among existing motion space examples, the motion-gradient space, usually written as ∇u, is most popular. So, the motion gradient space is used for the convenience of delivering the following analysis.

The interest of studying the potential, underlying connection between different penalisation functions (in regularisation schemes) are originated from the remarks of Black and Anandan (1996), who stated that: in terms of constructing the spatial coherence term within the least-squares formulation (corresponding to using the Tikhonov penalisation), the variation in the motion gradient field is assumed to be Gaussian. Later, along with the popularity of the L1-norm based regularisation strategies, the close connection between the L1-norm penaliser and the Laplacian distribution has been revealed. See Section 3.3.2,

and (Tibshirani, 1996), where detailed mathematic interpretations are provided between the L1-norm based regularisation and the Laplacian distribution. Furthermore, very recently, Krähenbühl and Koltun (2012) successfully demonstrated how to approximate well-known penalisation functions (in regularisation terms), including the L1-norm, the Lorentzian, and the generalised Charbonnier penalisation functions, with a limited number of exponential distribution kernels. This largely promotes the overall understanding about the close relationships between different penalisers based motion regularisation terms.

So, it can be understood that different motion-distribution priors (MDPs), in spite of corresponding to distinct formulations of penalisation functions, can be unified into generalised, probability distribution models. For example, the L1-norm, the Lorentzian, and the generalised Charbonnier penalisation functions can be all associated into the Gaussian Scale Mixture (GSM) model (Krähenbühl and Koltun, 2012); and, the classic Tikhonov and the popular L1-norm based penalisers both belong to the generalised Gaussian distribution family.

3.6 A New Generalised Motion Regularisation Formulation

The standard formulation for smoothness based regularisation terms is usually written as:

$$E_{reg}(\mathbf{v}) = \int_{\Omega} \omega \Psi(D\mathbf{v}) dx dy, \tag{3.21}$$

where $D\mathbf{v}$ denotes performing a particular order of differential operation on the field $\mathbf{v}$. It is common to see ω as a flexible parameter within the formulation. This will lead to spatially adaptive motion integration regions. For the choice of the derivative operator D, the f.o.d based operator has enjoyed popularity in OF estimation for a long time. Recently, many second-order based regularisation terms have begun to take the place. These methods assume that the underlying motion obeying piecewise affine transformations, divergence-free

motion patterns, etc., which can be depicted by second-order signal correlations. Specifically, different mixtures of second-order motion derivatives have been used (see examples of (Trobin et al., 2008) and (Kadri-Harouna et al., 2013)). In practice, the function Ψ can be the quadratic penalisation function $\Phi(z) = z^2$ (corresponding to the well-known Tikhonov penaliser) or one of the robust functions as listed in Section 3.2.1. Notice, D and Ψ are usually pre-selected and fixed for an individual motion estimation task.

For sparsity priors based approaches, the regularisation term is usually defined as:

$$E_{reg}(\mathbf{v}) = |W\mathbf{v}|, \tag{3.22}$$

where W represents the wavelet (or similar) transformation that is applied to the field $\mathbf{v}$. So, $W\mathbf{v}$ will be the transformed coefficients. The total variation or the L1-norm operator $|\cdot|$ is used to penalise large magnitudes of the coefficients. It is commonplace to choose the L1 norm, since it is known to induce sparse solutions (Shen and Wu, 2010, Nawaz et al., 2011, Han et al., 2011). In these types of regularisation schemes, the only task is to define W, which usually has a constant form, i.e. the basis for the corresponding transformation is pre-defined.

The way of formulating the aforementioned regularisation terms satisfy the common understanding about how to construct a particular regularisation scheme. However, these formulations of standard regularities, except for obviously allowing spatially adaptive, motion integration regions, seems to indulge the "may-be-improper" custom that relies on uniform motion spaces as well as globally fixed MDPs in OF estimation.

Considering the problem of selecting MDPs, local distribution statistics in a target motion space can be different from the global distribution. This will make it inappropriate to set a constant MDP over the whole motion field. In addition, existing regularisation approaches usually, empirically choose non-adaptive motion spaces, as well as MDPs. Looking through the literature, there are few regularisation strategies that enable the spatially adaptive selection of a motion space or a MDP, or are concerned with automatically deciding optimal choices for both of the two elements. To the best of our knowledge,

47

Myronenko and Song (2009) present the only work (for image registration) that simultaneously estimates motion flows as well as an optimal motion space (corresponding to treating the operator D in Equation (3.21) as an unknown). Regarding the MDP, it is only recent that two works appear in the literature, one treating the penalisation function Ψ (in Equation (3.21)) as an unknown parameter, and the other work (Chantas et al., 2014) suggesting a way to spatially adjust the variance parameter for the employed Gaussian-distribution prior based regularity.

According to the previous analysis, it can be understood that the existing, standard ways of formulating regularisation terms are not appropriate. To be more concrete, the symbols D and Ψ (both defined in Equation (3.21)), as well as W (in Equation (3.22)) should all be allowed to vary spatially, rather than being fixed over the whole motion field.

Motivated by the previous understanding, the three elements, namely the motion-supporter range, the motion space/domain, and the associated motion-distribution model, have been integrated together. This has lead to the following, newly generalised motion regularisation term,

$$E_{reg}^0(\mathbf{v}, L, \tilde{\Psi}) = \int_\Omega \omega \tilde{\Psi}(L\mathbf{v}) dx dy, \tag{3.23}$$

The immediate aim of designing such a formulation is to clearly assign roles to all of the three key tasks, that are most important in the construction of a particular regularisation scheme. Explicitly, ω decides the motion integration region that can be adaptively set for each pixel involved; $L\mathbf{v}$ represents the operation that transforms the field $\mathbf{v}$ into a desired domain or space, such as the motion-gradient domain ($\nabla\mathbf{v}$), second-order derivatives (s.o.ds) based domains (e.g., $\nabla^2\mathbf{v}$ or the Laplacian of $\mathbf{v}$), a wavelet-transform induced domain, etc. Since L can be any derivative operator or wave-form basis function, the smoothness based and the sparsity based regularisation schemes are both included in the new regularisation formulation. $\tilde{\Psi}(\cdot)$ is designed in a way that enables spatially flexible choices of penalisation functions, so as to respect local statistical distributions in the considered motion space (determined by L).

Compared with existing regularisation terms, the novelty of the newly generalised regularity is reflected at the following aspects:

48

- L and $\tilde{\Psi}$, respectively associated with the motion-space transformer and the smoothness penalisation function, can be both treated as unknowns.

- The three parameters/operations ω, L and $\tilde{\Psi}$ are all allowed to be adapted spatially.

Furthermore, since the spatially adjustable function $\tilde{\Psi}(\cdot)$ is just a qualitative definition, it is not clear what kinds of penalisation functions can be chosen from, and how they can be varied spatially. The following introduces two candidate functions that quantitatively formulate $\tilde{\Psi}(\cdot)$, so as to facilitate the algorithm realisation for practical requirements.

1. Generalised Gaussian distribution (GGD) based $\tilde{\Psi}$ (Mohammad-Djafari, 2012, Cho et al., 2010):

$$\tilde{\Psi}(L\mathbf{v}, \eta, \sigma) = \frac{|L\mathbf{v}|^{\eta}}{\sigma}, \tag{3.24}$$

The hyperparameters $\eta \in (0, 2]$ and $\sigma \in (0, \infty)$ are the shape parameters of the local latent distributions in the considered motion space. Specifically, η determines the peakiness of the distribution, and σ the width of the distribution. The larger the value of η, the flatter the distribution shape; the smaller η, the more peaked the shape. Due to the design, both convex and non-convex penalisers are taken into account in the regularisation term; and, the L1-norm and the Tikhonov penalisers are particular cases for the penalisation function Ψ, with $\eta = 1$ and 2 respectively.

The relationship between the above function of $\tilde{\Psi}$ and a GGD based pdf is given by,

$$\begin{aligned} \mathrm{pdf}(L\mathbf{v}, \eta, \sigma) &= e^{-\frac{|L\mathbf{v}|^{\eta}}{\sigma}} \\ -\log(\mathrm{pdf}(L\mathbf{v}, \eta, \sigma)) &= \tilde{\Psi}(L\mathbf{v}, \eta, \sigma) \end{aligned} \tag{3.25}$$

That is why $\tilde{\Psi}$, as defined in Equation (3.24), determines a GGD based MDP for the regularisation term in Equation (3.23).

Note that the space of the considered penalisation function in Equation (3.24) spans a spectrum of feasible density functions, varying from sharper non-convex functions to smoother convex functions, so, the associated regularisation term can spatially, adaptively enforce different degrees of sparsities based MDPs.

2. Gaussian Scale Mixture (GSM) (Wainwright and Simoncelli, 2000) based $\tilde{\Psi}$:

$$\tilde{\Psi}(L\mathbf{v}; \tilde{\Omega}) = \sum_{k=1}^{K} \tilde{\omega}_k(-\log(\mathcal{G}_k(L\mathbf{v}, 0, \tilde{\sigma}_k))),$$

$$\mathcal{G}(x, 0, \sigma) = e^{-\frac{x^2}{2\sigma^2}}$$

$$(3.26)$$

in which $\tilde{\Omega} = \tilde{\omega}_1, \tilde{\omega}_2, ..., \tilde{\omega}_K$ are the weights of the GSM model, and $\tilde{\sigma}_k$ are the variance parameter for the corresponding Gaussian kernel $\mathcal{G}_k$. GSMs can also model a wide range of distributions ranging from Gaussians to heavy-tailed ones. Taking $\tilde{\Psi}(L\mathbf{v}; \tilde{\Omega})$ into the generalised regularisation term leads to a GSM based prior adaptive regularisation term.

In practice, still other generalised, probabilistic distribution models can be considered to construct the function of $\tilde{\Psi}$, only with different numbers and types of hyperparameters to be involved. It is out of the scope of this thesis to find out all potential models.

Thanks to the generalised penalisation function $\tilde{\Psi}$, a very straightforward way has been enabled for the construction of new regularisation terms. More importantly, the generalised regularisation term is by itself a novel regularity. This is due to the fact that the model parameters, associated with the MDPs and/or motion spaces, are treated as unknowns in the generalised formulation. Therefore, the new regularisation scheme essentially allows the local adaptation of MDPs or motion spaces, or the optimal estimation of a global MDP or motion space. This is quite different from traditional regularisation approaches, which only estimate the flow field $\mathbf{v}$, and treat the motion-model parameters as pre-known.

3.7 Research Gaps and Promising Directions

According to the literature investigation and the generalised regularisation formation, some research gaps and promising directions have been detected as below:

3.7.1 In Relation to Motion Integration Regions

Within non-local regularisation approaches, a popular way of choosing motion supporters (see the definition of motion supporters in Section 3.2.2), or deciding a motion integration region, is the Bilateral filter based strategy (Sun, Roth and Black, 2010, Werlberger et al., 2010, Krähenbühl and Koltun, 2012). These approaches basically measure pairwise pixel/feature affinities according to pixel proximity, colour similarity etc. By expanding the motion integration range into larger and non-local scales, it becomes more difficult to accurately estimate mutual affinities between pixels. For some challenging situations, such as recovering motions for nearby objects with similar colour or appearance, the Bilateral-filter based strategies can easily lead to erroneous measurements.

3.7.2 In Relation to Regularisation Models

In non-local regularisation or layer based regularisation approaches, the selection of motion-prior models is never a trivial problem, due to the fact that a large region possibly contains multiple motions, or non-rigid motions, etc. Both learning and learning-free processes based regularisation approaches have been designed in the literature, that embed high-level knowledge about the underlying motion coherence or sparsity characteristics. Compared with learning-free regularisation approaches, learning based methods can theoretically enable the modelling of more complex motion priors. Namely, learning processes based approaches have more flexibilities in choosing motion domains, and capturing more complex statistics about the motion-domain distributions. However, training based strategies have not been able to outperform the more simple learning-free regularities, which can be seen from the evaluation results as presented in the Middlebury Benchmark[8]. One reason is that the challenge of learning "typical" flow patterns may not be feasible. This is true, given that different image structures and unknown object deformations may give rise to a multitude of motion patterns with little resemblance between motion fields from different datasets (Wedel et al., 2009). In addition, for the case of learning sparsity-prior models, dictionary learning methods usually aim to

---

[8]http://vision.middlebury.edu/flow/data/

train a universal and over-complete dictionary , in order to represent various motion-signal structures. However, sparse decomposition over a highly redundant dictionary is potentially unstable and tends to generate visual artifacts (Dong et al., 2011). By considering these factors, this research work will focus on learning-free regularisation approaches.

There are several aspects to consider regarding the possibility of designing new motion prior models.

1. Adaptive MDPs for the considered motion domain.

- For the case of motion coherence based regularisation, usually a fixed model of SoPMI is applied. For example, given a s.o.d based motion domain: if the L1-norm based penalisation is applied, such as in the works of Sun, Roth and Black (2010) and Werlberger et al. (2010), the s.o.ds of the flow field will be enforced to be piece-wise constant. This regularisation scheme is able to recover piece-wise affine motions in an image. However, if non-rigid motions exist in the same image, the estimated motion flow will be over-smoothed. In the case of exploiting the Tikhonov (quadratic) based function, the second-order motion derivatives will be encouraged to obey the Gaussian distribution, which has been proved suitable for recovering non-rigid motions. The quadratic penalisation combined with the s.o.d based motion domain is commonly seen in the image registration field , which is so called the curvature based regularisation approach (Fischer and Modersitzki, 2003). The rationale behind this is, by enforcing the s.o.ds to be sparse, high curvature motion is penalised, the quadratic penalisation however allows relatively denser distribution of motion curvatures. So, within a complex motion field that includes both rigid and non-rigid motions, more flexible, and possibly spatially adaptive MDPs are needed.

- Existing sparsity priors based regularisation approaches (Shen and Wu, 2010, Chen et al., 2012, Nawaz et al., 2011, Han et al., 2011), all assume a pre-defined sparsity-inducing function (usually L1-norm based). This, however, ignores the fact that the degree of the structure sparsity may vary spatially within a single motion field, and hence leads to

over-sparsified solutions.

To summarise, the assumption of a fixed MDP model does not respect the variant motion statistics within different datasets, or even within the same image. Therefore, an adaptive scheme needs to be designed.

2. Seeking new motion spaces or domains for non-local regularisation approaches. Most of the existing regularisation approaches only utilise local range signal correlations or deviations within the flow field, which is a fact in not only the derivative based, but also the wavelet based regularisation groups. In another word, the motion coherence or the sparsity prior based regularities only take into account motion correlation/deviation relationships between or among adjacent pixels in a local neighbourhood. Recently, the works of Sun, Roth and Black (2010) and Werlberger et al. (2010) open a door for the non-local range of motion signal correlations. However, these regularisation approaches are confined to non-local pairwise affinities in the original motion field (i.e., zeroth-order motion derivative space), but have not stretched out to explore non-local correlations in other motion spaces, such as the projected spaces after gradient, curvature or wavelet transformations. Through observations, the phenomenon of information repetition has been found not only in the original motion field, but also in the f.o.d based motion space. This thus suggests a direction for researchers to investigate new motion domains for non-local regularisation.

3.8 Conclusions

In this chapter, a comprehensive literature survey has been provided on motion-regularisation approaches in OF algorithms. Then, essential issues that are of crucial importance to the construction of modern regularisation strategies, including the motion-supporter selection, choosing motion spaces, and making assumptions about the statistics in the selected motion space, have been identified and discussed in detail. After that, the underlying connections between existing motion regularisation strategies are extracted. This finally leads to a new generalised formulation for state-of-the-art regularisation schemes. Due

to the indepth literature investigation, research gaps have been revealed and promising research directions are suggested. Based on this, the following chapter delves into the topic of non-local regularisation approaches, and Chapter 8 is to investigate adaptive MDP based regularisation strategies.

Chapter 4

A Non-Local Regularisation Approach based on Oriented Geodesic Distance

Optic-flow (OF) estimation needs spatial coherence regularisation, due to local image noise and the well-known aperture problem. More recently, OF local-region regularisation has been extended to larger or non-local regions of regularisation, to further deal with the aperture problem. After a careful literature review, it has been determined that the criteria used for deciding the degree of non-local pairwise motion coherence can be further improved. For this reason, an oriented geodesic distance (OGD) based motion regularisation scheme is proposed. The approach is particularly useful in difficult motion estimation situations, such as recovering motions for nearby objects with similar appearance, and reducing errors in object boundary regions. Experimental results, comparing to leading-edge non-local regularisation schemes, have confirmed the superior performance of the proposed approach.

The chapter is organised as follows. In Section 1, the motivations and the contributions are stated. Section 2 introduces traditional OF techniques, and also reviews the most relevant and related OF regularisation schemes in the literature. Section 3 defines the proposed OGD scheme, and provides the explanation for the designed non-local OF approach. Section 4 first details the

implementation of the proposed approach, then delivers the testing results and makes comparisons of the proposed approach with the state-of-the-art related non-local OF approaches. In Section 5, future work based on the proposed method is briefly discussed, and finally the conclusions of this chapter are presented.

4.1 Motivations and Contributions

Existing motion coherence schemes for non-local OF estimation can be characterized into image-driven and motion-driven approaches: The **image-driven category** compares low-level image feature statistics, such as pixel position similarity, colour similarity, etc, to decide on motion-coherence degrees for the associated pixels. The **motion-driven category** enforces spatial motion smoothness according to motion similarities. Therefore, motion-driven schemes need accurate initialisation of motion estimation, which has to be made in advance through image-driven based OF approaches. For the image-driven category, the most advanced regularisation schemes borrow ideas from the image segmentation and image denoising fields. More clearly, pairwise feature (pixel) similarity measurements are exploited. For example, Sun, Roth and Black (2010) and Werlberger et al. (2010) take advantage of Bilateral Filter (BF) based schemes to guide motion-coherence decisions, and Ren (2008) exploits the Intervening Contour (IC), which is first introduced in the field of image segmentation, to measure feature similarities and decide pairwise motion-coherence degrees.

Although improvements have been achieved in OF motion estimation, by expanding the motion supporter region (MSR) into larger and non-local scales, more efforts are still needed to improve the accuracy of measuring pairwise or groupwise pixel affinities. The task becomes more challenging for non-local range pixels, considering that enlarging MSR can proportionally increase the chance of connecting/grouping pixels from different motion surfaces. In existing non-local regularisation schemes such as (Sun, Roth and Black, 2010, Werlberger et al., 2010), the employed pairwise affinity measurers only take into account the information carried by the associated pixels, and ignore the

transition information between the pixels. While, the ignored information between the transition path of two pixels can be vital for correctly measuring the affinity between the pixels. This has thus inspired the utilisation of GD based non-local OF regularisation approaches, since GD provides a natural way to include the transition information that can help to accurately decide pairwise pixel affinities. In this regard, an OGD based non-local OF approach is proposed.

The novelty of the proposed approach has two aspects: 1) to the best of this author's knowledge, this is the first time that GD based measurements have been proposed into non-local OF regularisation schemes. In this regard, an isotropic geodesic distance (IGD) based OF regularisation scheme is suggested. 2) A new orientated GD path strategy is designed and integrated into the previous IGD based scheme, in order to help the IGD based OF estimation to combat the "motion leaking" problem (see definition in Figure 4.1), which is a commonly detected problem near object boundary regions.

4.2   Literature Background

4.2.1   A Short Review on Classic OF Algorithms

Existing OF approaches all rely on the temporal conservation of some image elements, e.g. conservation of pixel intensity or gradient values, etc. Commonly used elements are pixels, since they can be easily extracted and lead to dense measurements. See Equation 2.1, as introduced in Section 2.1.1, describing the most straightforward data-conservation constraint.

This chapter aims to propose a novel regularisation approach. Before that, the two classic regularisation schemes are briefly reviewed as below:
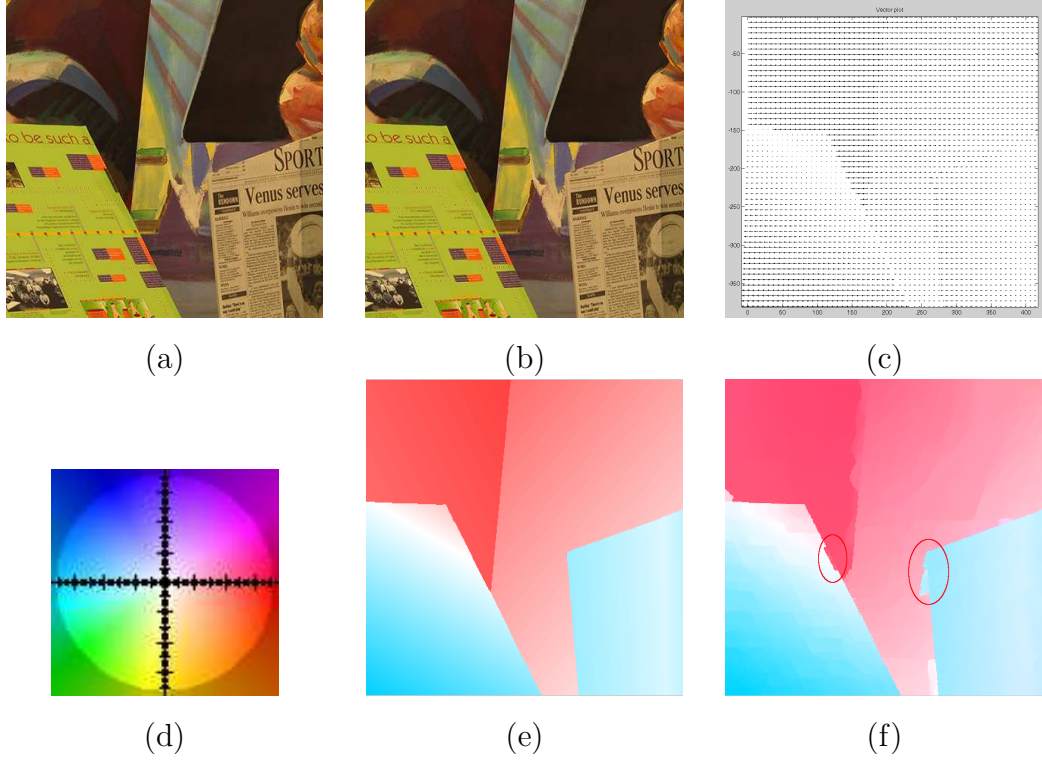
Fig. 4.1: (a) and (b) show two adjacent Venus images. (c) shows the ground-truth OF vector field between the two images. (d) presents the standard colour map that encodes flow vectors. Compared with the colour-coded ground-truth OF map in (e), (f) illustrates the "motion leaking" problem that takes place where the foreground object (in the Venus images) and the background share similar colours on some boundary regions (marked by red circles).

- The Horn-Schunck (HS) based scheme (Horn and Schunck, 1981):

$$E_S(u,v) = E_S(u) + E_S(v),$$

$$E_S(u) = \sum_{(x,y)\in\Omega} \{(\alpha_{rht}^{(x,y)}|u_{x,y} - u_{x+1,y}|^2 + \alpha_{bott}^{(x,y)}|u_{x,y} - u_{x,y+1}|^2 +$$

$$\alpha_{left}^{(x,y)}|u_{x,y} - u_{x-1,y}|^2 + \alpha_{top}^{(x,y)}|u_{x,y} - u_{x,y-1}|^2)\} \qquad (4.1)$$

$$E_S(v) = \sum_{(x,y)\in\Omega} \{(\alpha_{rht}^{(x,y)}|v_{x,y} - v_{x+1,y}|^2 + \alpha_{bott}^{(x,y)}|v_{x,y} - v_{x,y+1}|^2 +$$

$$\alpha_{left}^{(x,y)}|v_{x,y} - v_{x-1,y}|^2 + \alpha_{top}^{(x,y)}|v_{x,y} - v_{x,y-1}|^2)\}$$

where (x, y) denotes a space point in the 2-D image domain, and (u, v) is the velocity field of the considered image. $(u_{x,y} - u_{x+1,y})$ is the discrete, forward first-order derivative (f.o.d) of u with respect to x, and the rest are defined in a similar manner. Noticing that the parameter

58

$\alpha^{(x,y)}$ is not contained in the original HS OF algorithm. It is included here to recall the theories of both isotropic and anisotropic smoothness regularities that will be discussed in the next section. By exploiting constant values of $\alpha^{(x,y)}$ everywhere in the OF field, such as in (Horn and Schunck, 1981), it will lead to local isotropic regularisation schemes; In the case of spatially adjusting the value of $\alpha^{(x,y)}$, varying strengths of smoothness constraints will be enabled between the associated flow vectors, resulting in anisotropic regularities (as introduced in Section 3.2.2). The regulariser (defined in Equation (4.1)) assumes that, within the whole flow field Ω, directly adjacent pixels share similar motions, which is achieved by minimising $E_S(u, v)$.

- Lucas-Kanade (LK) based scheme (Lucas and Kanade, 1981):

$$E(\mathbf{v_x}) = \sum_{\mathbf{x'} \in N} \omega_{\mathbf{x},\mathbf{x'}} (I_2(\mathbf{x'} + \mathbf{v}) - I_1(\mathbf{x'}))^2 \tag{4.2}$$

where I_1 and I_2 denote the intensity functions of two consecutive image frames. The vector symbols $\mathbf{v} = (u, v)^T$, and $\mathbf{x} = (x, y)^T$ are used for the simplicity of the formulation. The LK method assumes that the unknown OF vectors are constant or similar within a local neighborhood N centered at a pixel $\mathbf{x}$. The weight value $\omega_{\mathbf{x},\mathbf{x'}}$ determines how much the motion of the centered pixel $\mathbf{x}$ should be similar to its neighbour $\mathbf{x'}$.

4.2.2 Most Related Works

Depending on how the weight value $\alpha^{\mathbf{x}}$ or $\omega_{\mathbf{x},\mathbf{x'}}$ (both introduced in the last section) are defined, isotropic and anisotropic motion regularities have been proposed in the OF literature. For example, the Gaussian filter based approach assumes that the influence of neighbouring pixels is isotropically and proportional to pairwise pixel proximities. The Bilateral Filter (BL) based approach assumes the regularisation weight is determined by the proximity and colour affinity of pairwise pixels, as defined below (Sun, Roth and Black, 2010, Werlberger et al., 2010):

$$\omega_{\mathbf{x},\mathbf{x'}} = \exp\{-\frac{|\mathbf{x} - \mathbf{x'}|^2}{2\sigma_1} - \frac{|I(\mathbf{x}) - I(\mathbf{x'})|^2}{2\sigma_2}\} \tag{4.3}$$

where σ_1 and σ_2 are the distance variance and the colour variance of adjacent pixels around $\mathbf{x}$. By this means, motion influences of neighboring pixels that are likely to belong to a different surface can be suppressed. Another anisotropic example is the structure-adaptive based work, proposed by Nagel and Enkelmann (1986), where larger weights are biasedly assigned to neighbouring pixels that locate along the same local contour, and reduced coherence is assigned between edge pixels and their non-edge neighbours. This is useful when little evidence can suggest which of the non-edge neighbours are located on the same surface with the edge pixel.

Considering that local regularisation based OF approaches take only a small neighbourhood into account, motion estimation can be troublesome in poorly textured regions and with edge features, i.e, those image regions find no matching and ambiguous motion vectors respectively. Hence, regularisation schemes applied within expanded regions have been proposed in modern OF approaches (Sun, Roth and Black, 2010, Werlberger et al., 2010), allowing motion cohesion to be established between pixel pairs within non-local image ranges.

Currently, the non-local OF regularisation schemes with top performances rely on the non-local BL (Sun, Roth and Black, 2010, Werlberger et al., 2010), and the Intervening Contour (IC) techniques. The IC based scheme uses a straight line connecting the studied pair of features (pixels), and assigns lower affinity to a pixel pair if the associated IC path passes more high gradient pixels. The potential weaknesses of these methods can be summarised: 1) BL considers no transition information between pixels, so the BL based scheme may enforce an improper motion coherence between two pixels that locate on different objects or motion regions. As shown in Figure 4.2, pixels A and B have different motion ground truth, however, BL based regularisation schemes will assign A and B large affinity considering their large proximity and intensity similarities. 2) IC does not allow a curved path, and can thus assign an inaccurate affinity for a pixel pair such as C and D, although they share the same motion region on the chair (also see Figure 4.2).
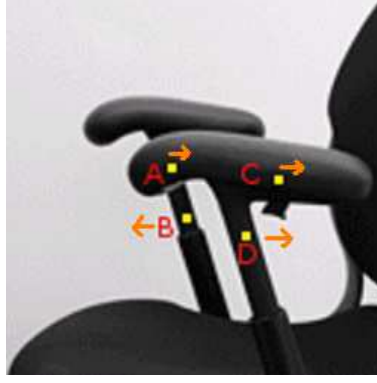
Fig. 4.2: An image of a rotating chair, with human marked ground truth motion vectors for selected pixels.

4.3 Oriented Geodesic Distance based Optic-Flow

4.3.1 Geodesic Distance based Regularisation

Given two pixels inside an image frame, the geodesic distance in-between is the shortest weighted length (according to a given metric), over all paths that connect the pair of pixels (Cohen and Kimmel, 1997, Benmansour and Cohen, 2011). The definition of GD algorithms is repeated here for convenience. Specifically, a GD path between pixels $\mathbf{x}_1$ and $\mathbf{x}_n$ can be written as:

$$D_1(\mathbf{x}_1, \mathbf{x}_n) = \min_{P(\mathbf{x}_1, \mathbf{x}_n)} \sum_{i=1}^{n} |\nabla I(\mathbf{x}_i)|^2 \qquad (4.4)$$

or

$$D_2(\mathbf{x}_1, \mathbf{x}_n) = \min_{P(\mathbf{x}_1, \mathbf{x}_n)} \sum_{i=1}^{n-1} |I(\mathbf{x}_{i+1}) - I(\mathbf{x}_i)| \qquad (4.5)$$

where $P(\mathbf{x}_1, \mathbf{x}_n)$ is an arbitrary, parameterised discrete path, with n pixels given by $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$. $\mathbf{x}_1$ denotes the start pixel and $\mathbf{x}_n$ the end pixel. The quantity $|\nabla I(\mathbf{x}_i)|$ (inside D_1) is a finite difference approximation of the image gradient at location $\mathbf{x}_i$. The quadratic term of the image gradient penalises a GD path going across image edges. For the second GD path example D_2, $|I(\mathbf{x}_{i+1}) - I(\mathbf{x}_i)|$ calculates the oriented first-order derivative of image intensity between the pixels $(\mathbf{x}_i, \mathbf{x}_{i+1})$ (Gulshan et al., 2010).

Considering the same pixel samples as shown in Figure 4.2, a low pairwise affinity will be determined for pixels A and B by GD based measurements, because of the large gradient region between A and B. However, other affinity measurers such as BL or the Gaussian filter based ones will provide improper measurements, due to the ignorance of image transition information. In addition, a GD based affinity measurement is more suitable for points C and D, considering that GD allows curved paths that can circumvent obstacles between C and D. This is an important advantage of GD over its competitors such as IC based methods.

Based on these observations, a GD based pairwise affinity measurement is suggested as a new motion regularisation weight,

$$\omega_{\mathbf{x}_1, \mathbf{x}_n} = D_{gd}(\mathbf{x}_1, \mathbf{x}_n) \tag{4.6}$$

$$D_{gd}(\mathbf{x}_1, \mathbf{x}_n) = \min_{P(\mathbf{x}_1, \mathbf{x}_n)} \sum_{i=1}^{n-1} |I(\mathbf{x}_i) - I(\mathbf{x}_1)| \tag{4.7}$$

This newly defined regularisation weight $\omega_{\mathbf{x}_1, \mathbf{x}_n}$ is to replace the weight value α as defined in Equation (2.3) and $\omega_{\mathbf{x}, \mathbf{x}'}$ in Equation (4.2). By doing so, a new MSR based regularisation term can be enabled in the OF estimation process.

It is worth noting that although the D_1 type of GD path has been successfully exploited in image segmentation tasks, such as in the work of Gulshan et al. (2010), it intrinsically forbids to contain pixels with large gradient magnitudes. So, for edge and texture types of seed pixels, a considerable rate of valuable motion supporters are excluded. A GD path defined as D_2 might follow a path with gradually different colour from the start pixel. Therefore, the selected GD scheme is defined as $|I(\mathbf{x}_i) - I(\mathbf{x}_1)|$, which is the colour difference between any pixel on the path and the start pixel. By setting $\mathbf{x} = \mathbf{x}_1$ and $\mathbf{x}' = \mathbf{x}_n$, the new, isotropic geodesic distance (IGD) based motion regularisation weight is thus:

$$\omega_{\mathbf{x}, \mathbf{x}'} = D_{gd}(\mathbf{x}, \mathbf{x}') \tag{4.8}$$

4.3.2 Oriented Geodesic Distance based Regularisation

In addition to the newly introduced, GD based motion regularisation scheme, an improved version that is based on an oriented GD measurement is further

proposed:

$$\omega_{\mathbf{x},\mathbf{x}'}^{ogd} = D_{ogd}(\mathbf{x}, \mathbf{x}') \tag{4.9}$$

where $D_{ogd}(\mathbf{x}, \mathbf{x}')$ represents the oriented geodesic distance between the two pixels. Detailed definition for $D_{ogd}(\mathbf{x}, \mathbf{x}')$ will be later provided in this same section.

The introduction of the OGD based scheme is, on one hand, aimed to conquer the intrinsic problem that is associated with the IGD based scheme, i.e. the "motion leaking" problem. Figure 4.1 illustrates the "motion leaking" problem taking place on object boundaries, where there are ambiguous colour/intensity distributions between the foreground and background regions. For example, the right boundary of the book and the left boundary region of the sports newspaper (both marked with red circles). On the other hand, the OGD scheme can also relieve a difficult problem in OF estimation: recovering motions for boundaries within textureless objects. Rather than exploiting the methods, such as that proposed by Ren, using flow-drive schemes to select a relatively higher affinity region from the two side regions (separated by the boundary). A much safer way is to enforce smoothness along the contour itself. This logic can be further supported by the structure-adaptive anisotropic scheme, and the contour based motion regularisation method (Liu et al., 2006).

$$D_{ogd}(\mathbf{x}_1, \mathbf{x}_n) = \min_{P(\mathbf{x}_1,\mathbf{x}_n)} \sum_{i=1}^{n-1} (\sqrt{|I(\mathbf{x}_i) - I(\mathbf{x}_1)|^2 + \tilde{\eta}(\mathbf{x}_1)(T(\mathbf{x}_i) \cdot (\mathbf{x}_{i+1} - \mathbf{x}_i))}$$

$$\tag{4.10}$$

The proposed formulation of the OGD is defined as in Equation (4.10). The first quantity in the equation is the same as that introduced in Equation (4.7). The second quantity is a structure-tensor based, direction-similarity term, which is measured by the dot product of the tensor directions of two subsequent pixels. The structure tensor is a field of symmetric positive matrices that encode the local anisotropy of an image. It was initially introduced for corner detection (Forstner, 1986). The second term in Equation (4.10) thus enables GD paths to travel along local and non-local structures. Importantly, a new weight $\eta(\mathbf{x})$ is introduced that decides where it is necessary to strongly enforce the structure-oriented GD searching, i.e. encouraging edge features to

have motion coherence along the same contour. This is achieved by taking $\eta(\mathbf{x})$ as the large eigen-value associated with the start point, which is a measurement of how confidently the start point is locating on object boundaries. The eigen-value field is normalised into the range of $[0,1]$, written as $\tilde{\eta}(\mathbf{x})$. Also, the two quantity terms have been normalised. So, for high confidence edge features, their $\tilde{\eta}(\mathbf{x})$ values will be approximately 1, and the two terms in the equation will be evenly enforced. For homogeneous regions, the colour-similarity based IGD searching is used. Figure 4.3 provides an intuitive illustration for the motion supporter regions (MSRs) (for two pre-selected pixels), that are selected by the proposed OGD scheme and the intensity similarity based GD strategy. Notice, for the same boundary pixel on the edge of the book, the MSRs that are respectively decided by OGD and GD based schemes have quite different shapes. While, for the pixel in the textureless region of the book, the selected MSRs are similar.

Two sampled pixels   MSRs selected by GD   MSRs selected by OGD

Fig. 4.3: Examples of the MSRs that are respectively selected by the intensity-similarity based IGD and the OGD schemes for the same two pixels (as denoted in the first picture).

4.4   Unified Optic-Flow Formulation

The proposed non-local OF approach consists of several components: a data term, a smoothness term and, the introduced adaptive weight that enables spatially variant MSR priors. The unified objective function is defined as:

$$E_{NL}(\mathbf{x}, \mathbf{v}) = \sum_{\mathbf{x} \in \Omega} R(I_2(\mathbf{x}+\mathbf{v}) - I_1(\mathbf{x})) + \sum_{\mathbf{x} \in \Omega} \sum_{\mathbf{x}' \in N_{ogd}} \{\omega_{\mathbf{x},\mathbf{x}'}^{ogd} S(\mathbf{v}(\mathbf{x}) - \mathbf{v}(\mathbf{x}'))\} \quad (4.11)$$

64

$R(z)$ and $S(z)$ are robust penalisation functions as introduced by Black and Anandan (1996), Sun, Roth and Black (2010), which penalise $1st$-order data deviations. Different penalisers can be chosen, such as the quadratic penalty z^2, the Lorentzian penalty $log(1 + \frac{z^2}{2\sigma^2})$, and the total-variation (TV) penalty $|z|$. The latter two are penalisers less sensitive to data or smoothness term violations, such as that caused by large illumination changes and occlusions. In this work, both the data and the smoothness terms choose the TV penalty due to its robustness to outliers. A much very detailed reference to robust penalisation functions is the work of Black and Anandan (1996).

Note that at the heart of the proposed regularisation approach essentially is the OGD based non-local MSR strategy, which is encoded in the weighting parameter $\omega_{\mathbf{x},\mathbf{x}'}^{ogd}$ that is associated with the smoothness term.

The formulation of Equation (4.11) is straightforward to understand. However, because of the TV terms, the energy functional is now non-convex, and needs special treatments to achieve minimisation. Methods such as Graduated Non-Convexity (GNC) can be adopted (Black and Anandan, 1996). Details are included in the next section.

4.4.1 Implementation Associated Problems

One major problem with the OF differential formation is the estimation of large displacements. Existing strategies for dealing with large motion estimation include, the decoupled minimisation without warping (Steinbrucker et al., 2009), the graph-cut method (Freedman and Turek, 2005), the multi-resolution based method, etc. The decoupled minimisation has implementation simplicity but lacks sub-pixel accuracy. The graph-cut based implementation is relatively computational expensive. The proposed method is embedded within the multi-resolution approach. This is a common way to overcome the difficulty of large motion estimation by creating an image pyramid (constructed from original images). Within such a framework, the main components of displacements $\mathbf{v}$ are first estimated at a coarse resolution. Then, the estimation is progressively refined while going down the pyramid level.

In order to deal with the non-convex energy functional minimisation, a GNC scheme is exploited and embedded within the multi-resolution framework. The general idea is to take the non-convex object function and construct a corresponding convex approximation (Black and Anandan, 1996). The minimisation is then processed in a continuation method, that starts from minimising the constructed convex functions to minimising the targeted non-convex function. Successively better approximations of the true objective function can thus be steadily achieved. For a given non-convex objective function, the difficult part is how to construct the sequence of approximate convex functions. Further detail on this topic is available in (Black and Anandan, 1996, Blake and Zisserman, 1987).

In addition, the proposed OGD path-searching scheme is efficiently computed using the Fast Marching Method (Peyré et al., 2010). The computation complexity is $O(nlog(m))$, with n being the number of total pixels in the image field, and m the number of local and non-local neighbours that are checked for each pixel involved.

4.5 Performance Evaluation

The following datasets, with ground truth OF, are chosen for expriments: the Aloe, Venus, RubberWhale, Wood, and Dimetrodon image sequences (see Figure 4.4), all of which are extracted from the well-known Middlebury Benchmark [1]. Since the proposed algorithm is designed to estimate two-frame motion flows, only one image pair is randomly selected, from each image dataset in the experiments.

For performance comparison, the results from the top-performing non-local OF work of Sun, Roth and Black (2010). The two-frame versions of the corresponding datasets are taken as input. The standard Intensity Constancy Assumption (ICA) is adopted in the selected data term. Both the GD and BL based regularisation schemes use the same RGB colour space for the purpose of selecting motion supporter regions. In addition, for all the image pairs,

---

[1]http://vision.middlebury.edu/flow/data/

| Image | GT | Sun-NLOF | OGD |
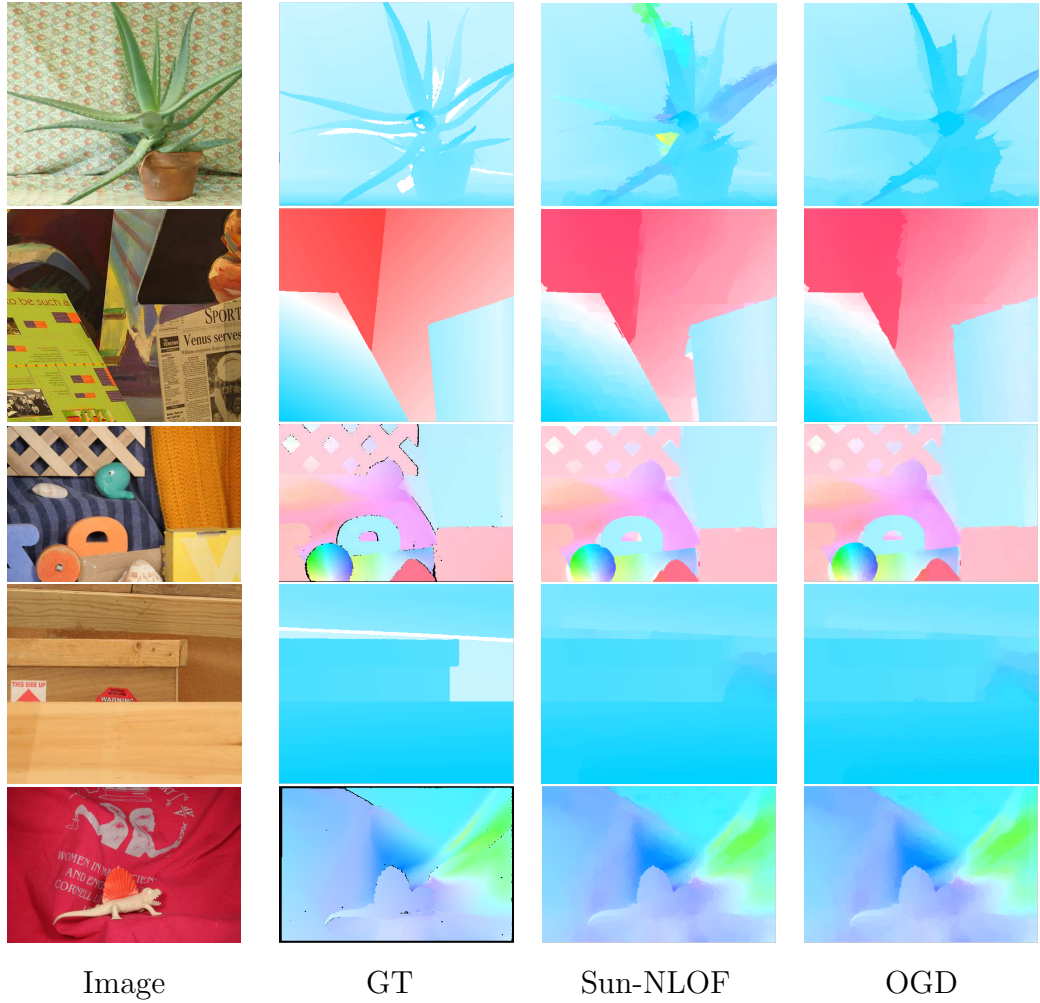
Fig. 4.4: Experimental results for the selected images. Actual flows are visualised using the standard colour coding. The ground truth flows (the second column), the results from Sun, Roth and Black (2010), denoted as Sun-NLOF (the third column), and our OGD results (the last column) are displayed.

the same structures of GNC and multi-resolution schemes are used. To be specific, the GNC stages take the first two stages for minimising the approximated energy functions, and the third stage is for the true objective function. Within the first two stages, the number of pyramid levels is two, and the last stage contains a five-level pyramid. Also for the purpose of fair comparison, a uniform number of non-local neighbours are used in the proposed approach as well as the approaches compared, specifically, 15×15 neighbours in the finest image layer, and gradually reduced numbers of neighbours in the remaining layers.

For an intuitive demonstration, Figure 4.4 displays the visualised colour maps for the estimated OF fields, the flow results of (Sun, Roth and Black, 2010) and the corresponding ground truth flow fields. Quantitative results are also presented, as listed in Table 4.1. In all cases, standard measurements of average angular errors (AAE) and average end point errors (EPE) (as used in (Barron et al., 1994)) are reported, except that some testing results are not available for the work of Ren (2008).

Table 4.1: Quantitative results

	OGD		Sun-NLOF		Ren-OF	
	AAE	EPE	AAE	EPE	AAE	EPE
Venus	3.093	0.231	3.500	0.251	3.930	0.260
RubberWhale	2.480	0.076	2.469	0.077	5.320	0.170
Dimetrodon	2.579	0.131	2.656	0.136	3.340	0.170
Aloe	0.686	2.187	1.932	2.705	NA	NA
Wood	0.384	3.554	0.388	3.584	NA	NA

For the five datasets being tested, the proposed OGD based approach has achieved better accuracies in four cases and one comparable result. The motion estimation for Aloe, RubberWhale and Wood images, with nearby similar-colour objects/regions, are challenging tasks. The Venus and RubberWhale images have ambiguous boundary regions between nearby objects. So, they are suitable to be included for testing whether the OGD based regularisation can achieve the expected outcome, i.e., solving or relieving the "motion leaking" problem. The Wood and the Dimetrodon images contain not only clustered similar-colour objects/regions, but also large amounts of textureless regions. So, these two datasets are to validate the robustness of the OGD based scheme for reliably grouping non-local ranges of motion supporters. In addition, images have rigid-motion flows, namely the Dimentrodon dataset, are also tested through experiments.

Particularly, the proposed method reports lower motion estimation errors in the Aloe and Wood images, where similar-colour objects are clustered together.

The "motion leaking" problem, which can be seen existing in the testing results of Sun, Roth and Black (2010) and Ren (2008), has been efficiently suppressed by the proposed OGD based regularisation approach. For the demonstration purpose, a close-up comparison for the three approaches is displayed in Figure 4.5. Notice that more crispy motion boundaries have been achieved by the proposed method. Some very ambiguous boundary regions are highlighted, as shown in Figure 4.5, by red circles that are superimposed on the colour-coded flow maps.

| Ren-OF | Sun-NLOF | OGD |
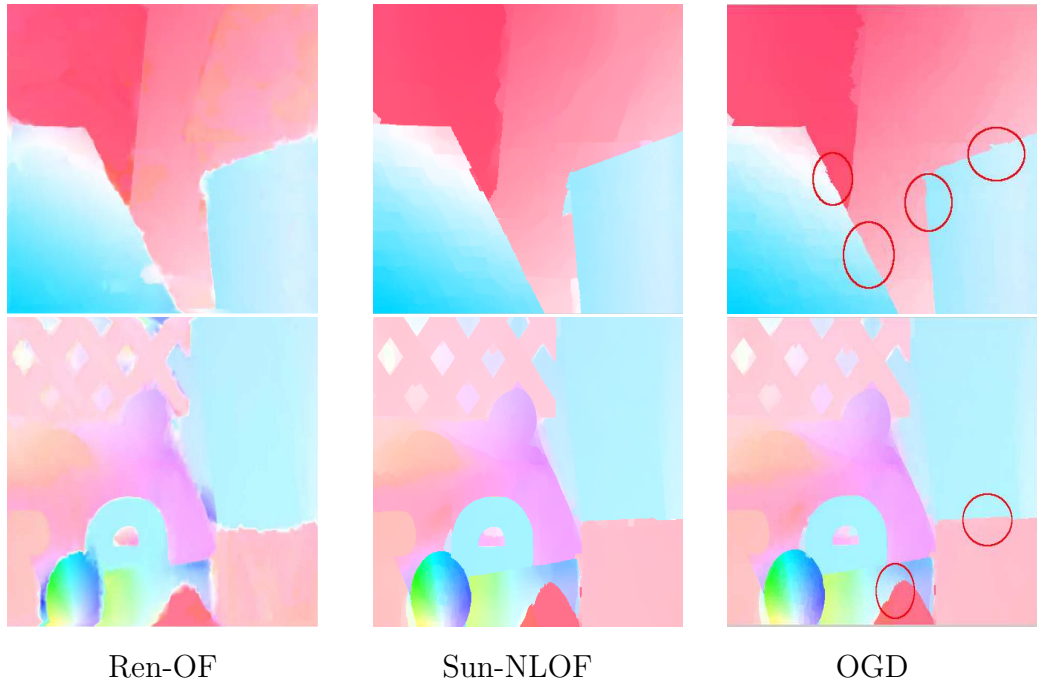
Fig. 4.5: Illustrating the relieved "motion leaking" problem. See the red circles that mark the corresponding regions where there are much crispy motion boundaries. The testing results according to the work of Ren (2008) are denoted as Ren-OF.

The quantitative results in Table 4.2 show the motion estimation accuracies of the OGD based approach and the method of Sun, Roth and Black (2010), for different regions in the RubberWhale image (see the manually segmented regions in Figure 4.6). By further analysing the statistics in Table 4.1 and Table 4.2, it is revealed that the OGD based method has enabled superior motion estimation accuracies in object boundary regions, textureless objects, and non-rigid motion situations. The improved accuracies in object boundary regions

confirm the effects of the proposed OGD strategy. It is noticed that OGD has a problem in estimating motions of occlusion regions. To analyse the reason, see the occlusion region, as illustrated in Figure 4.7. This type of occlusion regions are usually not corresponding to object boundaries. However, the structure tensor based strategy, inside the OGD scheme, cannot distinguish the "false boundaries" with real ones. Those "false boundaries" possess no meaningful motion information, so, the motion regularisation enforced along the "false boundary" contours will result in erroneous motion vectors. Considering that the occlusion region detection and the associated motion estimation is by itself very challenging, future research on the topic will be conducted.
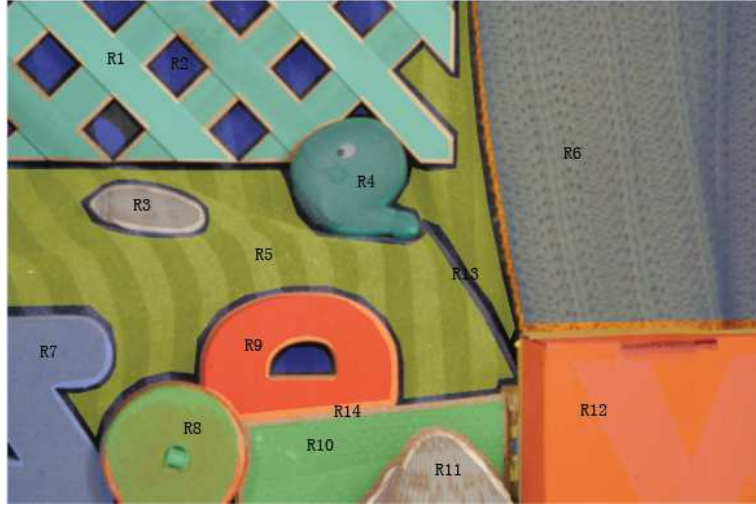
Fig. 4.6: Illustrating some of the manually segmented regions inside the RubberWhale image.

Table 4.2: AAE and EPE for regions in RubberWhale

	OGD		Sun-NLOF	
	AAE	EPE	AAE	EPE
Rotation (R8,10)	2.405	0.114	2.459	0.120
Aperture (R2)	8.613	0.204	9.099	0.215
Repetitive texture (R6,10)	0.896	0.031	0.828	0.029
Average texture (R3,11)	1.640	0.044	1.433	0.039
Sparse texture (R4,8)	2.201	0.066	1.951	0.061
Textureless (R1,7,9,12)	1.504	0.052	1.656	0.057
Non-rigid (R5)	2.094	0.064	2.224	0.067
Occlusion (R13)	9.099	0.230	8.637	0.220
Boundary (R14)	3.566	0.118	3.968	0.130

(a) Venus image
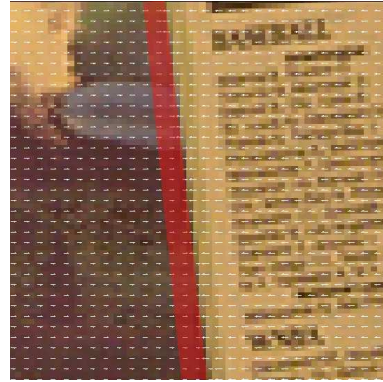
(b) Enlarged ground truth OF

Fig. 4.7: (a) illustrates the occlusion region (marked with red colour) in the Venus image. The ground truth flow for a cropped occlusion region is enlarged and displayed in (b).

4.6   Conclusions

In this chapter, geodesic distance based non-local regularisation schemes have been investigated for the first time. Particularly, an oriented geodesic distance (OGD) based, non-local range pixel grouping, or image segmentation algorithm has been designed. This OGD algorithm has been exploited in the development of a new, non-local regularisation approach for OF estimation. Through experiments on five real datasets, the proposed non-local OF approach has been demonstrated particular useful in dealing with two challenging situations: accurately recovering motion flows near object boundary regions and estimating motions for nearby similar-appearance objects, where the existing competing approaches have inferior performance.

Chapter 5

Cell Population Tracking: Background and Literature Suvey

The purpose of this chapter is two-fold. First, the chapter introduces the background of computer-vision based cell tracking approaches. In this regard, the motivations and challenges about tracking dense cell populations over phase-contrast image sequences are described. Then, three main categories of existing cell tracking approaches are explained, including the detection/segmentation and association based, model based, and motion based cell tracking approaches. Secondly, the chapter presents a literature survey on key approaches and systems, that have been recently proposed and successfully applied to tracking dense and relatively dense cell populations. After that, quantitative comparisons are made for the related systems and frameworks. Based on the investigations, promising research directions are suggested.

5.1 Introduction

Cell dynamics is a field of intense current research in which researchers pursue improved comprehension of fundamental processes in cellular and developmental biology. Cell behaviour such as migration (translocation), proliferation

(growth and division) and differentiation, and apoptosis[1] play a central role in many fundamental biological processes.

As the process of increasing cell numbers by reproducing cells themselves (see the illustration in Figure 5.1), cell proliferation is most evident during embryonic development, where many phases of growth and division generate the cell mass of the developing organism. After the adult organism is formed, many cells, such as in the case of neurons and muscles, permanently possess the ability of growth. And for other cells, they retain the option of growing again temporarily to replace cells that are lost either through the normal processes of wear and tear, or through damage (Berridge, 2012). The process of cell differentiation performs specialised functions, expressing cell-type-specific genes, to generate different types of cells, such as bone cells, muscle cells, neurons, etc., that shape different organ developments. Failures in the cell differentiation process can cause defects, such as brain malformations. Apoptosis, i.e. a process that controls cell death as illustrated in Figure 5.2, is also an important event in development, when unwanted cells are pruned during organogenesis[2]. For instance, the apoptosis of cells located in-between the fingers and toes allow for their separation. In an adult human body, apoptosis continues to play a vital role during the normal turnover of cells. For example, the human body spawns several million new cells every second, and correspondingly several million cells must die for the body to remain cell number constant. Problems with the regulation of apoptosis have been implicated in a number of diseases, such as cancer, a disease which is often characterised by too little apoptosis (Berridge, 2012).

All of these processes require the orchestrated movement of cells in particular directions to specific locations, and changes of the cellular shape will also often determine the fate of the cell. So, understanding the mechanisms of cell migration and morphology is thus one of the important goals of biomedical research.

---

[1]Apoptosis, also called programmed cell death, is a process in which cells play an active role in their own death, distinct from another form of cell death that is uncontrolled by cells themselves.

[2]Organogenesis: (in embryology) the formation and differentiation of organs and organ systems during embryonic development. In humans, the period extends from approximately the end of the second week through the eighth week of gestation.
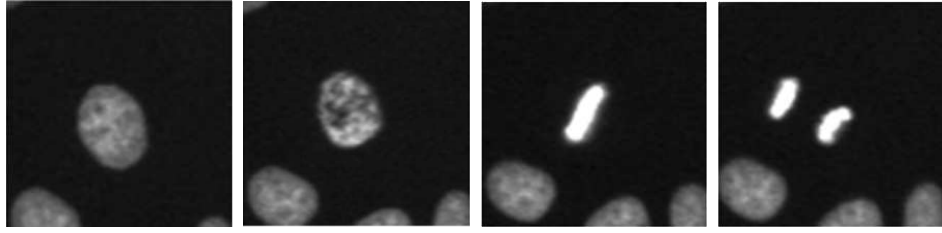
Fig. 5.1: The four images illustrate a cell division process (from left to right), where the mother cell exhibits colour and appearance changes and it is gradually divided into two child cells. The images are taken from the paper of Li et al. (2010).
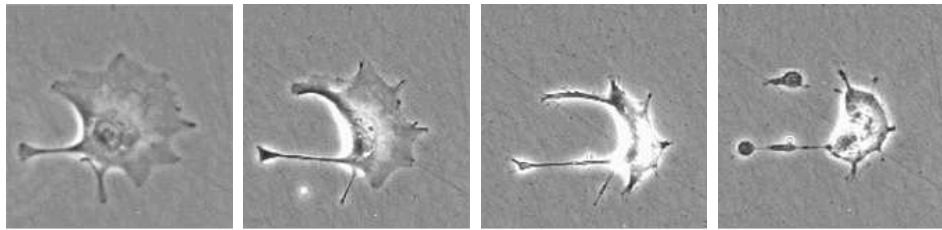
Fig. 5.2: Illustration of a cell apoptosis process, where the cell gradually changes appearance and loses dynamic. The images are taken from the website [3] of the Cell Migration lab at the University of Reading.

5.1.1 The Need for Automated Approaches

The need for automated approaches is motivated by the fact that cellular behaviour of migration, proliferation and apoptosis involve considerable numbers of cells. Reliable analysis of cell dynamics requires the use of non-fluorescence microscopy[4], such as phase-contrast microscopy[5] for a long period of time (usually several days). This will routinely produce thousands of images with low signal-to-noise ratios. Considering the high processing requirement of manually tracking cells, i.e. marking the location and/or the shape of each individual cells in every image frame, it will be tedious and time-consuming. Moreover,

---

[4]A fluorescence microscope is an optical microscope, used to study specimens that contain materials, which can be made to fluoresce. In this technique, high intensity light illuminates the sample and excites fluorescence species in the sample, which then emit light of a longer wavelength.

[5]Phase-contrast microscopy is an optical microscopy illumination technique in which small phase shifts in the light passing through a transparent specimen are converted into amplitude or contrast changes in the image.

the results often lack reproducibility, even if the tracking work is performed by experts, which is largely due to unavoidable user bias and subjectivity.
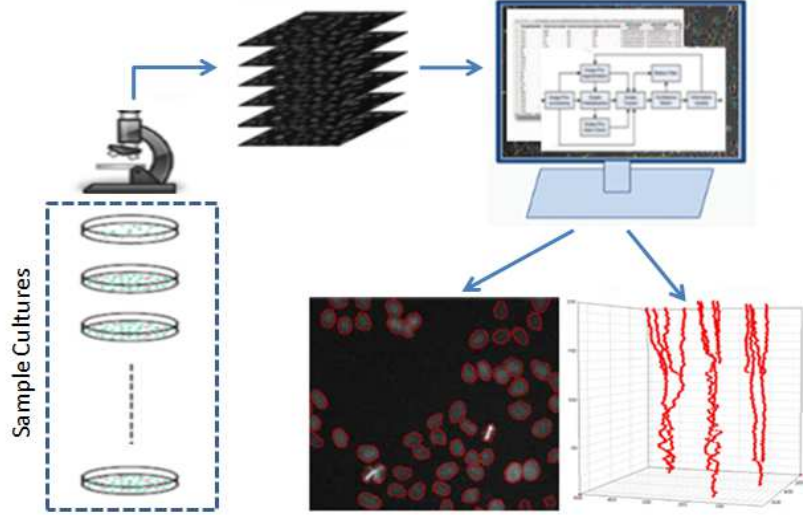
Fig. 5.3: An intuitive system that flows from the cellular image recording process to the procedure that outputs cell tracking results. The cellular images and the cells' trajectory map are taken from the paper of Li et al. (2010)

To illustrate the main procedures involved in cell tracking, Figure 5.3 presents a flow diagram, depicting the following procedures:

- Culturing *in vitro* cells in planes.

- Recording cellular images using microscopy devices.

- Visually tracking cells by automated systems that take cellular images as inputs.

- Outputting cell tracking results, such as cell trajectories and cell deformation information, over the time-lapse image sequence.

5.1.2 Challenges of Vision-based Cell Tracking

The current survey is very interested in the problem of tracking cells in highly concentrated datasets, especially those taken from a phase-contrast time-lapse

microscopy. Figure 5.4 displays two samples of phase-contrast cellular images. Due to the halo regions surrounding cell membranes, phase-contrast cellular images are usually challenging to segment. Within dense cell populations, cells can be clustered, and have frequent interactions, thus correct segmentation is non-trivial. Since a large variety of cell shapes appear, this poses great difficulties in cell shape modelling as well as segmentation. In addition, spatially variant cell densities also lead to spatially varying motion patterns, so, motion fields of dense cell populations are extremely challenging to estimate.
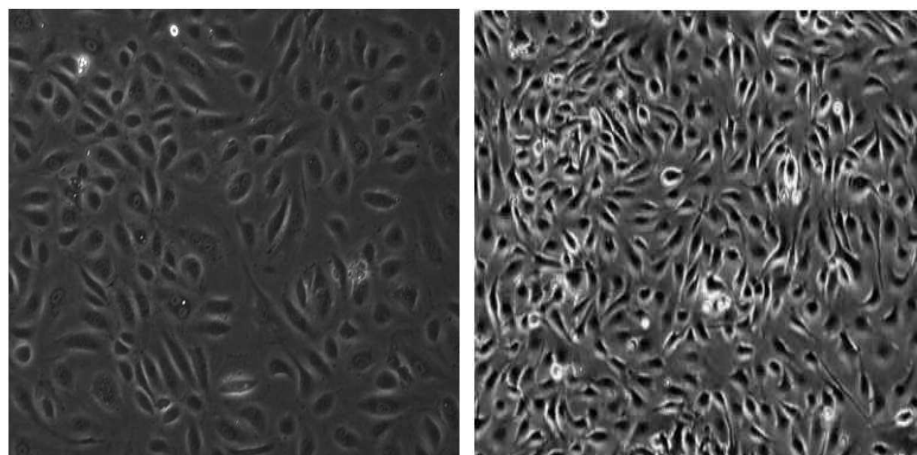
Fig. 5.4: Illustrating ill-defined cell boundaries, diversity of cell shapes, cell interactions or cell partially overlapping, and cell occlusion situations.

A more complete summary of challenging problems that are associated with dense population tracking (over phase-contrast image sequences) are listed as below:

- Ill-defined cell boundaries - the low contrast between cells and the background, due to the lack of optic focus, depending on cell positions with respect to the focal plane. The first image in Figure 5.5 illustrates ill-defined, blurred cell boundaries.

- Similar intensity distributions between the foreground and background image regions. See the middle image in Figure 5.5.

- Different cell sizes, and diversity of cell shapes (see Figure 5.4).

- Considering that image sequences can be taken under different time
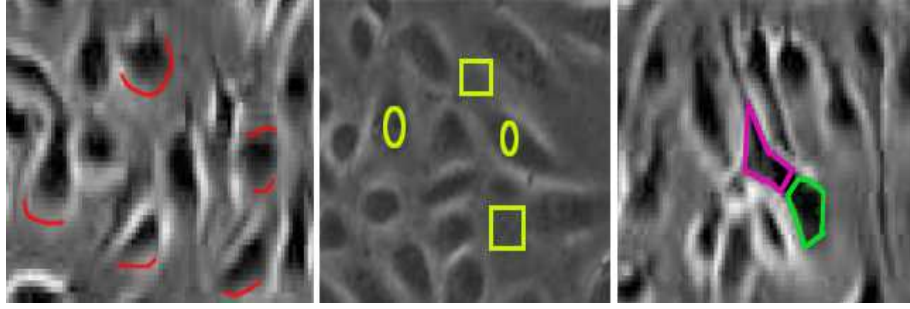
Fig. 5.5: The first image illustrates ill-defined cell boundaries that are manually marked by red contours. In the middle image, ellipse and rectangular regions are respectively intercellular regions and image background regions, which demonstrate that similar intensities can exist between image foreground and background regions. The third image displays two partially overlapped cells with their interface regions merged.

frequencies, large cell-shape deformations and/or cell displacements between successive frames are possible.

- Nearby cells interact with each other, where the cell boundaries in-between can become invisible. The third image in Figure 5.5 displays two partially overlapped cells.

- Figure 5.6 illustrates some marked elongated cells, which can easily lead to image over-segmentation.

- Spatially variant motion patterns in cellular images of dense populations.

- Cell division, newly incoming cells, and cells exiting from the image, and cell debris. These events pose difficulties to correct cell-to-cell association. See Figure 5.1 for cell division and Figure 5.2 with a cell death event.

- Cells partially or fully occluded in the environment. This also poses the cell association problem.
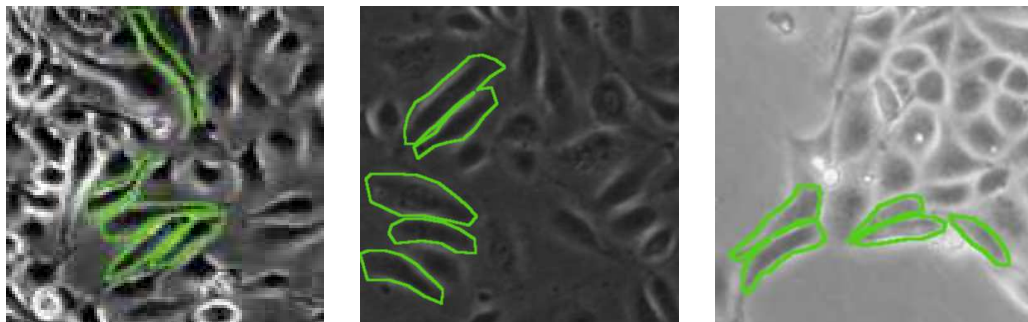
Fig. 5.6: Elongated cells, marked with the green colour.

5.2   Categories of Cell Tracking Methodologies

The goal of this section is to present a comprehensive literature survey on the topic of cell tracking. Firstly, existing approaches are divided into three categories: detection/segmentation and association based (also referred to as segmentation-association based), model-evolution based, and motion based cell tracking approaches. In this regard, the basic ideas of the three categories of approaches are briefly explained, in order to facilitate the upcoming detailed analysis about the relevant and related cell tracking approaches and frameworks in Section 5.3. Specially, reviews on recently proposed works that concern the tracking of relatively dense cell populations are focused upon. After that, a qualitative comparison is presented for the most related cell tracking approaches or systems.

5.2.1   Segmentation-Association based Tracking

In a typical detection/segmentation and association based approach, the tracking task includes two major steps: the detection/segmentation step and the association step. First, the feature detection or image segmentation process tries to identify individual cells in each frame within a given image sequence. According to specific microscopy techniques that are exploited for recording cellular images, and different types of cells to be tracked, distinguishable cellular features, such as colour and boundary properties can be exploited in the process of cell detection/segmentation, that aims to separate foreground cells

79

from image background regions, and also segment cells from each other in the same image field. Some methods satisfying these purposes include global/local thresholding approaches, peak detection methods, watershed segmentation algorithms (Vincent and Soille, 1991), etc. After separate cell regions are detected/segmented in each frame, the second task is to associate cells across image frames, i.e. matching cells between successive frames. To achieve this purpose, the pre-chosen cell properties, usually based on cell appearance and position information, can be employed.

- For measuring cell appearance properties, cell boundary and interior-region characteristics can be considered: 1) a straightforward way for describing cell boundary characteristics may take into account perimeter and/or curvature statistics of the boundaries. More accurately, the Fourier shape descriptor, firstly proposed by Cosgriff (1960), can also be employed in the representation of a closed cell boundary curve. 2) Regional attributes, extracted from relatively simple information such as cell area size and region compactness, or more complex measurements of the colour distribution in pre-segmented cell regions and cell texture statistics, can all be quantified for cells' similarity comparisons.

- In cell-association strategies that concern cell position similarities, the most simple method associates a cell in one frame with the spatially nearest cell in the subsequent frame. To avoid cell-matching ambiguities in the case of multiple cells close to each other, some advanced strategies take into account the neighbourhood relationship of a cell with its surrounding nearby cells, for the purpose of cell association. For example, in the work of Li et al. (2010), the variabilities of the spatial distribution of neighbouring nuclei are measured and compared, including the nearby nuclei identities surrounding the target nuclei, and the distances between the target nuclei and each of its neighbours. Except for exploiting spatial position similarities, the temporal context coherence of cell trajectories can also be considered in the cell association and tracking process. For example, Li et al. (2008) suggest to re-connect broken cell trajectory segments by overseeing the entire tracking history.

5.2.2 Model Evolution based Tracking

In model evolution based cell approaches, mathematical tools have been used to model cell appearances and/or cell motion characteristics, based on high level assumptions. Different from detection/segmentation based approaches that separately perform cell detection/segmentation and cell association processes, model based approaches aim to track cells by jointly updating cell location and region information frame by frame. In this category of approaches, the cell detection/segmentation is only required in the first frame, and the tracking results in one frame are used to initialise the tracking process in the subsequent frame.

Some representative algorithms have been successfully applied to model cell appearances, and cell movement features. For example, the mean-shift algorithm (Cheng, 1995) is adopted in the work of Debeir et al. (2005), where cells are modelled as bright/dark intensity dots. Parametric active contours (first proposed by Kass et al. (1988)), modelling cell membranes (such as demonstrated by Zimmer et al. (2002) and Ray et al. (2002)), due to the deformable property of snake models. The level-set framework based active contours, also called non-parametric active contours, are employed in the works of Padfield et al. (2009) and Li et al. (2008) for modelling cell shapes. As for motion models based cell tracking approaches, well-known examples include Kalman filtering (Ray et al., 2002) and particle filtering (Kitagawa, 1987) techniques, both of which can model cell dynamic features, such as velocity, acceleration and rotation properties, that can be learned from cell behaviour over image sequences.

5.2.3 Motion Estimation based Tracking

In addition to the aforementioned two categories of cell tracking approaches, researchers recently proposed direct methods that track cells according to pre-estimated motion fields. This has led to the third category of cell tracking approaches, which are referred to as motion based approaches in this research work. Basically, the idea of motion based cell tracking includes the following steps:

- An image segmentation process is performed in the first frame for initially segmenting cell regions.

- Exploiting a selected motion estimation technique to establish pixel-to-pixel or object-object mapping between successive frames of an image sequence.

- Starting from the initial segmentation results, cell regions are associated and tracked from frame to frame, according to the pre-estimated pixel- or object-level translocation information in every frame.

At the heart of the motion based cell tracking approaches is image-based motion estimation. For example, in the work of A.J. Hand and MacNeil (2009), the OF algorithm is exploited to match image pixels between two consecutive frames, according to pixel intensity similarities and gradient similarities. The average velocity of a pre-marked cell region is estimated at the initial stage, and then the motion information of the cell region is used for frame-to-frame tracking. Another motion based work is proposed by Yang et al. (2008), where a variant of the Demons algorithm (Thirion, 1998) (commonly used in medical image registration), is adopted to align cell-nuclei regions between consecutive frames. According to the works (A.J. Hand and MacNeil, 2009, Yang et al., 2008), a major advantage of optic-flow or image-registration based tracking systems is that the actual image acquisition process will not significantly affect the tracking performance.

This category of approaches requires separate image segmentation processes (either only in the first frame, or in a selected number of frames), and cell association processes. So, by analogy, the motion estimation based approach can be considered as a special type of segmentation-association based approach. Specifically, the cell association process is performed "on the fly", using pre-segmented cell regions in one frame as initialisation for cell detection/segmentation in the next frame. Essentially, this process matches one cell to its spatially closest cell in the subsequent frame.

5.3 Related Cell Tracking Systems and Comparisons

5.3.1 Cell Population Tracking Systems

The system of Jurrus et al. (2009) tracks axon populations within image sequences that are acquired by serial block-face scanning electron microscopy (SBFSEM). Radial active contours are applied to jointly segmenting and tracking axons, slice by slice. An OF technique provides velocity estimation for each pixel in the image. The Kalman filtering technique is used for predicting and correcting axon positions from noisy measurements that are estimated by active contours and the optic flow algorithm. The system has been proved useful for tracking axons with regularly round shapes, and relatively simple motion dynamics.

Within the approach of Padfield et al. (2009), the cell segmentation and tracking tasks are accomplished by the non-parametric active contour algorithm (based on the level-set framework). Pre-defined cell shape and size constraints are integrated into the active contour model to prevent the level-set contour from leaking out from weak edges of cell nuclei. For detecting and tracking cell division events, Padfield et al. (2009) design a speed function that is coupled with the Fast Marching approach, in which the associated energy model has a cost function that adaptively switches from favouring and emphasising bright appearance nuclei to dark nuclei along the image sequence. By doing so, the approach is able to track cells across the four stages of a cell cycle.

Li et al. (2008) present an automated system for visually tracking cell population in vitro phase-contrast images. The quantification of cell migration, mitosis and apoptosis has been enabled. Their system consists of a non-parametric active contour based cell tracker, in conjunction with a carefully designed, adaptive Interacting Multiple Models (IMM) of motion filtering, and a module for spatiotemporal trajectory optimisation. The curve merging problem, associated with standard, non-parametric active contours, is handled by integrating topology constraints that permit level-set division, but prohibit merging.

Rodriguez et al. (2009) introduce a multiple object tracking framework for unstructured crowded scenes, such as exhibitions, railway stations, and biological cells. The tracking system involves two key techniques: the OF algorithm and a Correlated Topic Model (CTM). The computed OF field is used as low-level features for each tested video samples. The CTM is trained to model various crowd behaviour modalities, with the associated parameters estimated through a collection of learning processes. One main contribution of their work is the idea of exploiting high-level knowledge that is learned from crowd behaviour for the task of object tracking.

The work of Li et al. (2010) track multiple cell nuclei in fluorescence microscopy images. A seeded-watershed algorithm is used for nuclei image segmentation. The nuclei association process takes into account properties of cell intensity profiles and shapes, and the measurements of cells' spatial neighbourhood structures. Although nuclei in the tested dataset tend to have a relatively high rate of large displacement, i.e. the cell displacement between two consecutive frames is larger than the cell's diameter, many of those cells have been successfully tracked. This is due to the help of a spatial structure-similarity based cell-matching strategy, which cooperates with the appearance-similarity based matching. For detecting dividing and newly divided nuclei, the morphologic and appearance changes of the nuclei have been exploited. Specifically, a support vector machine (SVM) classifier, taking into account six manually selected features, namely, cell sizes, average and standard intensities of cell regions, compactness of cells, the short and long axis of cells, is adopted for distinguishing general nuclei and division cells.

Kanade et al. (2011) present a segmentation based dense cell population tracking system, which consists of multiple modules: 1) The microscopy image restoration process, restoring the "authentic" microscopy images and producing images without artifacts such as halo and shading (that usually appear in the phase-contrast microscopy images). 2) The cell segmentation process, conducted on the resulting cleaner images; 3) The cell tracking module, using multiple hypothesise for cell association. 4) The cell-division detection module. Similar features are first extracted for manually extracted, candidate cell-division sequences, each consisting of a sequence of image patches cropped around the same cell. A training procedure is then conducted on a set of

duplicated patch sequences that contain various rotated patterns of the cells. After that, the task of determining whether a birth event happens and when it happens is decided firstly by a HCRF (Hidden Conditional Random Fields) model and then a trained SVM classifier.

Dewan et al. (2011) proposed a segmentation based method for automated cell tracking. The framework consists of three main modules: detection, tracking, and trajectory recovery. 1) The detection module detects and localises nuclei of cells. A morphological top-hat filter is employed for illumination correction and the h-maxima transformation (Soille, 2003) for the purpose of nuclei segmentation. Then, an ellipse is fitted for localising the cell nucleus. 2) To find the correspondence between cells, the following measurements are considered, motion parameters, skewness and displacement measurements and cell topological features, including colour compatibility, degree of area overlap, and deformation. One contribution of (Dewan et al., 2011) is the utilisation of the features from both of cell motion and topology domains, to track cells in clusters. 3) As also agreed by Dewan et al. (2011), it is hard to avoid segmentation errors completely, in segmentation-driven methods. To deal with that, along with cell division detection and tracking, a trajectory recovery model that is based on template matching is added to the framework.

In the work of Chatterjee et al. (2013), human monocyte cells are tracked in a fluorescent microscopic video, by matching and linking of bipartite graphs. Since image segmentation is relatively easy for selected cellular images, the main contribution of Chatterjee et al. (2013) is the modelling of cell association as a maximum-cardinality-minimum-weight matching problem for a bipartite graph.

Kaakinen et al. (2014) developed a segmentation-association based method that tracks cell migration over phase-contrast image sequences. In this approach, cell detection is considered as finding the maximally stable extremal regions (MSERs) in the image. Then, a Kalman filtering based multi-object tracker is exploited to estimate the migration magnitude and direction of individual cells in confluent cell populations. The approach is claimed to have the advantage of computation efficiency.

A.J. Hand and MacNeil (2009) proposed an OF based cell tracking approach. The system works well for tracking cells over a phase-contrast image sequence that contains up to 10 frames, and it has achieved a tracking accuracy of 95%. Their approach can not directly detect and track newly entering or newly born cells. In order to evaluate the system performance at different cell densities and over a larger number of frames, the authors made experiments on artificial image datasets, with different levels of cell population densities. The tracking performance degrades as the density of the targeted cells and the number of tested frames increase.

Yang et al. (2008) also developed a motion based cell tracking approach. The motion estimation process is accomplished by a variant of the Demons algorithm (Thirion, 1998) (commonly used in image registration). Pre-segmented cell-nuclei regions are then associated between consecutive frames, according to the estimated motion information.

In the work of Jiang et al. (2010), a SIFT based, motion estimation and tracking scheme is proposed for automated cell motility analysis, in low-contrast differential interference contrast (DIC) datasets. First, SIFT points around live cells are detected, and then a structure locality preservation (SLP) (using the Laplacian Eigenmap) is exploited to track the SIFT feature points along successive frames of DIC videos. The cell segmentation task is accomplished by the level-set method. The motion vectors that are obtained from the SIFT feature matching are used to associate cells between consecutive frames. In a clustered cellular environment, although the proposed work has not expressive tracking accuracies, Jiang et al. (2010) proved that the proposed Laplacian-SIFT can significantly reduce the error rate of SIFT feature matching, in comparison with principal component analysis (PCA) based SIFT tracking.

5.3.2 Qualitative Comparisons

To better analyse the performance of the related works that focus on the tracking of cell populations, a qualitative comparison is presented in Table 5.1 and 5.2. Not all of the previously reviewed works have been included in the table, considering that the work of Chatterjee et al. (2013) is validated

on not sufficiently dense cell populations, (Yang et al., 2008) is designed for cell tracking in 3-D datasets, and the approach of Jiang et al. (2010) has not achieved high accuracies in tracking dense cell populations. Specifically, the performance comparisons include several aspects: cell population densities, shape deformation tracking, cell division tracking, tracking cells with relatively large displacement (referred to as "jumping cells" in this section), tracking existing and newly incoming cells, occlusion handling, training and device requirements, etc. Except for the percentage values in the "Tracking accuracy" row, other fractional values (formatted as n out of 100) denote the (roughly estimated) occurrence frequencies of the corresponding events, which give clues about the difficulty level of dealing with those events in the corresponding cell tracking systems.
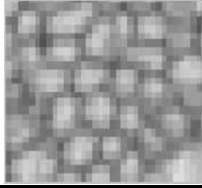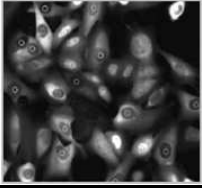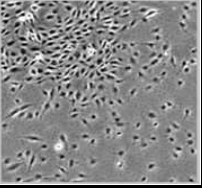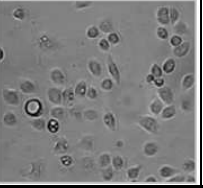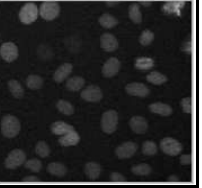
Sample					
Reference	Jurrus et al., 2009	Padfield et al., 2009	Li et al., 2008	A.J.Hand and MacNeil, 2009	Li et al., 2010
Microscopy	SBFSEM	Fluoresence	Phase-contrast	Phase-contrast	Fluoresence
Deformation level	small	NA	large	NA	small
Deformation tracking	yes	no	yes	yes	yes
Overlapping-cell tracking	no	no	yes, (3/100)	yes	yes, (15/100)
Jumping-cell tracking	no	no	yes, (5/100)	yes	yes, (25/100)
Mitosis tracking	no	yes, (10/100)	yes	no	yes, (15/100)
Occlusion-cell tracking	no	no	yes, (8/100)	no	no
Incoming-cell tracking	no	no	yes	no	yes
Apoptosis tracking	no	no	no	no	yes
Training requirement	no	no	yes	no	yes
Tracking accuracy	NA	NA	86.9%-92.5%	50.0%-60.0%	90.0%

Table 5.1: Qualitative comparisons of the related systems (part 1).
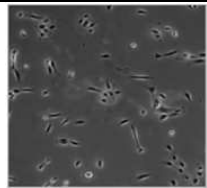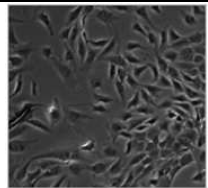
Sample				
Reference	Dewan et al., 2009	Kaakinen et al., 2013	Rodriguez et al., 2009	Kanade et al., 2011
Microscopy	Phase-contrast, fluoresence	Phase-contrast	Phase-contrast	Phase-contrast
Deformation level	small	medium	large	medium
Deformation tracking	yes	yes	no	yes
Overlapping-cell tracking	yes, (3/100)	yes, (3/100)	yes, (3/100)	yes, (25/100)
Jumping-cell tracking	yes,(5/100)	no	yes, (5/100)	yes, (5/100)
Mitosis tracking	yes	no	no	yes, (15/100)
Occlusion-cell tracking	no	yes, (4/100)	yes, (8/100)	no
Incoming-cell tracking	yes	no	yes	yes
Apoptosis tracking	no	no	no	no
Training requirement	no	no	yes	yes
Tracking accuracy	84.0%-87.0%	35.0%-65.0%	79.1%-89.0%	82.0%

Table 5.2: Qualitative comparisons of the related systems (part 2).

5.3.3 Discussion

According to the statistics in Table 5.1, discussions are presented in terms of the current trends and challenges in tracking dense cell populations.

1. **Over- and under-segmentation problems**: 1.1) segmentation based approaches have difficulties in segmenting cells within phase-contrast images. This can be caused by similar foreground and background intensity distributions, blurred cell boundaries, uneven cell intensities, cells textures (due to intracellular structures), image noise levels, etc. Fluorescence cellular images are relatively easier to segment due to the high image contrast. However, fluorescence-microscopy techniques usually require additional preparations such as staining, which may destroy the cells. Therefore, a large number of cellular datasets have been captured by non-fluorescent microscopies, such as the phase-contrast microscopy. In addition, there are following challenges:

- Segmentation based methods may encounter the over-segmentation problem in terms of cells with uneven intensities or abundant textures, and cells having a variety of shapes (especially elongated cells). These situations are commonly seen in dense cell populations.

- Since nearby cells may have overlapping regions, the boundaries in between will be ambiguous, which also poses difficulties for accurate segmentation. Kanade et al. (2011) proposed a method that enables the transformation from phase-contrast images to clearer images, i.e. images without artifacts. However, as long as there exist non-even interior cell regions, the overlapping of cells, and various shapes of cells, the problem of under- or over-segmentation cannot be avoidable in low-level image segmentation approaches.

1.2) Model based methods can potentially produce better estimates of cell morphologies (Li et al., 2008), due to the factor that high-level priors about cell shapes can be easily integrated in model based approaches.

2. One challenging problem, which is not well solved or often ignored in the existing works, is **the segmentation and tracking of elongated cells**, such

as those illustrated in Figure 5.6. Regions of elongated, long shape cells can be easily over-segmented and thus cause problem in segmentation-association based tracking methods. Long shape cells can also lead to frequent lost tracking in model based tracking systems. Existing model based approaches usually employ parametric/non-parametric active contours to model cell membranes and/or cell regions. For those active contours models that rely on the boundary based feature matching, elongated cells, due to the relatively low curvature boundaries, are difficult to track. Although some approaches also model cell regions, inferior tracking accuracies are usually achieved. This is because cell region models suffer the leaking problem, especially in the case of tracking dense cell populations within low contrast images. The suboptimal segmentation and tracking results of this type of cells can be easily accumulated over a number of frames, and lead to the tracking lost.

3. The problem of **tracking large displacement cells** exist in all of the three categorised approaches. When cells have large displacements between consecutive frames, segmentation-association based approaches encounter the problem of accurately associating correspondences among multiple cells. Model based approaches, because of the local minimisation problem, can not track a cell that has insufficient or no overlap with its translated cell. Motion based approaches, need to be embedded into a pyramid or multi-scale implementation framework, so as to track cells with relatively large displacements.

4. The task of **cell division detection and tracking** is gaining more attention. Cell division events are usually accompanied by large appearance changes. A large number of works employ training based processes, assuming that cells display particular appearance changes during the cell division process. For example, a most simple assumption is that cells will become more round in shape and abruptly display bright intensities. However, different types of cells, during cell divisions, may show large variances in appearances and behaviour (see Figure 5.7). Even within image sequences that contain the same type of cells, the time resolution of an image sequence has a significant influence on the cell division pattern. Therefore, training based approaches have a relatively weak generality in applications. In addition, considering that pre-training is required, those methods are usually not suitable for online event detection and tracking.
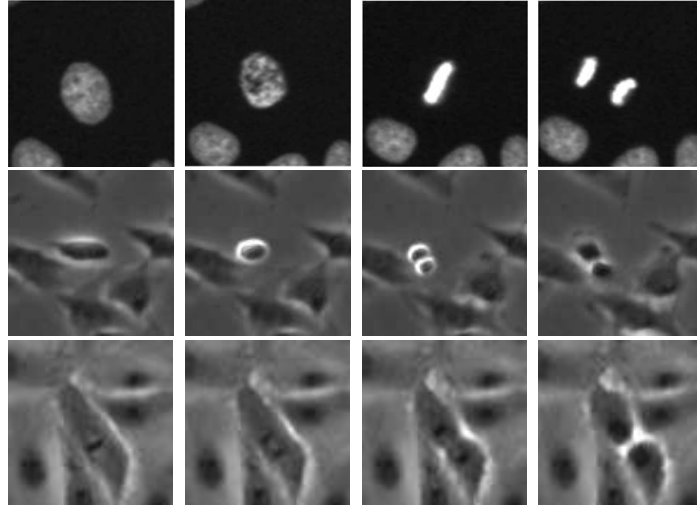
Fig. 5.7: Illustrating three samples of dividing cells that display differing patterns. Details of the sample images will be described in Section 6.3.

5. For tracking **newly incoming/entering cells**: in segmentation-association based frameworks, since all image frames in the image sequence are segmented, newly incoming cells can be naturally accommodated; motion based approaches and model based cannot directly deal with this issue, so, a separate cell segmentation or detection process is needed to examine newly appearing cells on image border regions.

6. For tracking **cells under occlusions**: Cells can be partially or totally occluded in the scene, due to the fact that they overlap heavily with the neighbours, or cells that exist the image field. Alternatively, cells may seem to be occluded when they have large appearance changes (usually during cell division or death). To deal with occluded cells, Li et al. (2008) exploit Kalman filtering based motion models that temporarily provide predicted positions for the cells. This, however, is risky under the condition of a dense cellular environment. Also, in the work of Li et al. (2008), a so-called track linker is proposed that aims to complete cell trajectories, by overseeing entire histories of cell trajectories, utilising spatiotemporal context information to re-connect broken cell trajectories.

7. **Training assistance** in tracking cell translocation: some existing works rely on training based methods to model cell behaviour. For example, Li et al. (2008) train four motion models. At run time, a cell's motion pattern

is associated to one of those four models, and the best-fitting motion model provides prediction information to help to track the associated cell, in the next one or two frames. Rodriguez et al. (2009) propose to learn motion direction priors of multi-modal crowd behaviour. However, training based processes are specifically designed for particular types of cells. Depending on the density level of the cell population, cell behaviours and cell interaction may be quite different.

5.4 Summaries and Prospective Research Directions

Segmentation-association based tracking approaches are relatively simple in their principle. These approaches are preferred for application to clear cellular datasets. If targeted images are recorded by phase-contrast microscopes, the cell detection/segmentation process will encounter particular difficulties. So, carefully designed, cell detection/segmentation strategies are usually required. In addition, non-trivial post-processing is also necessary to deal with the under- or over-segmentation problem, especially in the tracking of dense cell populations.

Model based approaches are in general not constrained by particular types of image datasets. They can potentially achieve relatively stable performance in segmenting cell shapes, and tracking cell populations with various densities. These advantages make model based approaches more suitable for tracking highly dense cell populations within phase-contrast image sequences. One weakness of model based approaches is the inability of dealing with newly incoming cells, which is due to the model initialisation problem. Since model based approaches suffer from the local minimisation problem, they are preferred to be applied to tracking cells with relatively small displacements between consecutive frames.

For motion based approaches, the main issue is the accumulated tracking error. In terms of tracking over a large number of frames, and with a high density of cell populations, the performance of motion based approaches is

inferior. Therefore, a straightforward suggestion is to combine motion based approaches with low- or high-level image segmentation tools. By this means, it is expected to provide refined cell segmentation results, so as to deal with error accumulation. One remarkable property of motion based approaches is that motion information in intercellular regions and boundary regions can be easily, simultaneously taken into account. This is potentially to provide valuable, complementary information, for model or segmentation based approaches, in segmenting and tracking partially overlapped cells, and also elongated cells.

When all of the merits and shortcomings are considered together, a promising direction is to construct a new cell tracking framework by drawing strengthes from different categories of approaches.

Another direction that deserves future research is the development of new strategies for detecting and tracking dividing cells. Based on the literature review, all of the three categorised approaches have difficulties in tracking division cells that have large appearance changes. Existing approaches usually rely on training based approaches. However, different types of cells can have large variability in their appearance and behaviour during the process of cell division. In addition, image capture frequencies can significantly affect cell division patterns. To be more concrete, cell appearance and morphology changes during the division process can have large differences with respect to the time resolution of the image sequence. These factors limit the application generality of existing, training based approaches.

5.5 Conclusions

The chapter first presents a background introduction on the topic of cell tracking, and then declares the the potential challenges in the design of automated cell tracking tools. After that, basic ideas of the three categorised cell tracking approaches are briefly explained, and a detailed literature survey is provided for the most related cell tracking approaches and systems. This has shed light on promising research directions for future work. Regarding this, a motion-occlusion analysis based approach will be proposed, in Chapter 6, that aims

to automatically detect cell division events. Chapter 7 will present a novel cell tracking framework, with the purpose of improving existing approaches in segmenting and tracking dense cell populations over phase-contrast image sequences.

Chapter 6

Cell Division Detection based on Motion Occlusion Analysis

The computer vision domain has seen increasing attention in the design of automated tools for cellular biology researchers. In addition to quantitative analysis on whole populations of cells, identification of cell division events is another important topic. In this chapter, a novel fully automated, cell-division detection approach is proposed. Differing from most of the existing approaches that exploit training based or image based segmentation methods, the main idea of the proposed approach is detecting cell divisions using a motion-based occlusion analysis process. Testing has been performed on different types of cellular datasets, including fluorescence images and phase-contrast datasets, and this has confirmed the effectiveness of the proposed method.

6.1   Existing Approaches for Mitosis Detection

Within the work of Padfield et al. (2009), dividing cells are assumed to have intensity changing from bright to dark. Division events are recognised and tracked within a unified framework that combines a level-set algorithm with a fast marching method, which has an associated cost function favoring objects with bright-to-dark changes. Nath et al. (2006) also apply a level-set method to the task of cell division detection and tracking. Level-set based methods,

are considered to be well suitable for identifying and tracking cell division, due to their flexible topology (Li et al., 2008, Nath et al., 2006). However, at least one important pre-condition should be satisfied: the mother cell in one image frame should have overlapping with its children cells in the subsequent frame. This pre-condition requires that the image sequence has a sufficiently high time resolution.

Li et al. (2008) consider that mitosis cells will turn into different appearances and shapes in the targeted datasets. A separate edge detector is used together with a simple mitosis pattern-recognition process for detecting cell divisions. A division detection method, contributed by Li et al. (2010), relies on a support vector machine (SVM) classifier, where six manually selected features (cell size, average intensity, standard deviation of intensity, compactness, long axis, and short axis) have been employed, for distinguishing general cells and division cells. Kanade et al. (2011) also apply a trained classifier to recognising cell division events. Huh and Chen (2011) present a probabilistic model, which is obtained by a learning process, for division detection. The aforementioned approaches have the limitation of being applied to particular datasets.

Quelhas et al. (2010) propose an OF based cell-division detection method. Their approach is based on speed difference of cell division events. Since Quelhas et al. (2010) assume that cell division as a much faster process, their approach cannot guarantee correct recognitions of division events when it is applied in a dataset with large variations of cell speeds, which is commonly seen in many cellular datasets.

Debeir et al. (2005) introduce a backward tracking strategy, i.e. tracking from the last frame to the first frame. In their approach, cell division events are detected when two cells gradually move towards the same location and finally merge together. The strategy is embedded within a mean-shift based cell tracking approach. As with model based, forward mitosis tracking methods, a requirement is that the cell displacement between two successive frames should not exceed the cell diameter.

For a much detailed literature survey on cell division detection, please refer to the work of Huh and Chen (2011).

6.2   Proposed Mitosis Detection Approach

At the heart of the proposed division detection approach is the assumption that where cell division happens, a group of pixels newly appearing/disappearing in one frame cannot be found in the adjacent frame. The most similar work that is also based on a motion estimation strategy for cell division detection is introduced by Quelhas et al. (2010). However, it is worth noting that neither particular changes about cell colour and shape, nor cell speed characteristics, such as that used by Quelhas et al. (2010), need to be assumed in the currently proposed approach.

In this section, the key techniques that make up the proposed approach are introduced. After that, the major components of the overall system are briefly explained.

6.2.1   Key Algorithms

Optic-Flow based Motion Estimation

OF estimation computes approximate motion fields for time-varying image sequences. Formulations of classic OF data and regularisation terms have been introduced in Sections 2.1 and 3.2. The basic idea is that, for each pixel (x, y) inside a frame I_1, a corresponding OF vector $(u_{x,y}, v_{x,y})$ is to be calculated, representing the pixel displacement between two consecutive frames I_1 and I_2. The data term measures how much the intensity information of the pixels is conserved during image transformation. The regularisation term constrains the smoothness of nearby OF vectors.

Within this work, the HS based OF algorithm has been chosen, which is featured by having both data conservation and motion smoothness terms, and leads to a dense OF field (u, v). Other motion estimation algorithms such as Lucas-Kanade (LK) (Lucas and Kanade, 1981) based ones rely on texture and corner features, and thus result in sparse flow fields. Demons based image registration methods (Thirion, 1998), which are firstly introduced for the estimation of non-rigid motion, lack smoothness assumptions, and so, usually

suffer from a relatively high rate of motion estimation noise.

One problem of the HS based optic flow is the estimation of large displacements. The chosen OF algorithm is thus embedded within the coarse-to-fine framework. Within such a framework, main components of the OF vectors are first estimated between down-sampled image pairs. Then, the higher resolution versions of the same images refine the flow estimation progressively.

Occlusion Detection

Given two subsequent image frames, occlusions that happen in motion estimation can be caused by the fact that pixels inside one of the two frames fail to be found inside the other. For example, when an object is shielded by other objects. Because of shade and/or illumination change, pixels that are mismatched between two frames also lead to motion occlusions. In this work, occlusions that are caused by newly incoming cells or cells exiting the image field will not be considered. This can be easily achieved by ignoring occlusion regions that are detected near image borders. Motion estimation errors due to illumination changes can be largely reduced by using a gradient-magnitude data conservation scheme.

The chosen occlusion detection method checks the mutual consistency between the forward and backward OF fields (Alvarez et al., 2007, Brox et al., 2004). Based on this scheme, an occlusion map O_1 is defined in this way: for a match between two pixels $(\mathbf{x}, \mathbf{x}')$ according to the backward OF field (the two pixels are respectively inside I_2 and I_1), if $\mathbf{x}'$ is not matched back to $\mathbf{x}$ using the forward flow field, the pixel $\mathbf{x}$ will be flagged as an occluded pixel, i.e. $O_1(\mathbf{x})$ will be set as 1. Flow vectors for off-grid pixels are obtained by interpolation. In addition, a second occlusion map O_2 is also constructed, defined as

$$O_2(\mathbf{x}) = \begin{cases} 0 & \text{if } \exists(\mathbf{x}' + \mathbf{v}'_{\mathbf{x}'}) \approx \mathbf{x}, \\ 1 & \text{otherwise.} \end{cases} \qquad (6.1)$$

where $\mathbf{v}'$ represents the forward OF field. The symbol $\approx$ denotes checking if $\mathbf{x}$ has been matched by any $\mathbf{x}'$ according to the forward flow.

In theory, O_1 and O_2 should have similar results, however, because of the

interpolation process required for O_1, there are some occlusions not detected inside O_1 but appear in O_2. Therefore, the final occlusion map is a combined binary field: $O = O_1 \cup O_2$. In practice, a gray-value occlusion map can be obtained, taking fractional values, ranging from 0 to 1, in both of the occlusion maps.

6.2.2   Major Functional Blocks

This work implements an automated cell-division detection system, which consists of four major components:

- Image based motion estimation. For every two consecutive frames within the image sequence, a forward-backward OF estimation, within the coarse-to-fine framework, is applied.

- Motion based occlusion detection. This process results in a binary or gray-value map that indicates where the occlusions happen.

- Morphology operations. Two morphology operations are conducted in order: a morphology erosion for removing scattered noises in the estimated occlusion map, and a morphology close for connecting nearby small occlusion regions.

- Cell division map output. This process calculates region centroids within the binary occlusion map. The regions that have the area size less then a threshold τ will be ignored. τ is manually or automatically selected according to the average cell size in the tested data. For a gray-value occlusion map, a simple peak detection method or the mean-shift algorithm (Cheng, 1995) can be used to detect the cell division positions.

6.3 Performance Evaluation

6.3.1 Testing Results

To demonstrate the validity of the proposed approach, experiments on several representative datasets have been conducted, including:

- Henrietta Lacks (Hela) - the donor of a cervical cancer cell dataset, captured under fluorescence microscopy. This dataset has been adopted in the work of Li et al. (2010).

- A phase-contrast image sequence of C2C12 myoblastic stem cells, which has been used by Huh and Chen (2011).

- A phase-contrast image sequence of wound healing (WH) cells (taken from (Nath et al., 2006)).

- Madin Darby Canine Kidney Epithelial cells (MDCK).

Due to different types of cells and microscopy techniques, cell division patterns differ from one dataset to another. Apart from that, for the three phase-contrast datasets, since the WH image sequence has a relatively high time resolution, the appearance and morphology changes of dividing cells in this sequence appear very slow from frame to frame.

In regard to the related parameters: 1) The morphology operation parameters. Both of the morphology operations take disk-shaped structuring elements with radius of 1 pixel; 2) The parameter τ, assuming the average cell size is a_c, can be set as $\gamma \times a_c$ with γ a positive fractional value less than 1; 3) The pyramid-level number l for the coarse-to-fine framework. With experiments, it is noticed that occlusions also happen in regions where cells have large displacements. Therefore, the value l should be accordingly adjusted for different cellular datasets.

The cell-division detection results returned by the proposed approach are compared against the ground truth, i.e., the manually annotated data by an expert. Three values are recorded for each image sequence, namely, the total number

of dividing cells that are manually annotated, the correct number of automatically detected division events (for each dividing cell, multiple detections on the same cell are counted just once), the number of non-division cells that are detected by the proposed method.

Figure 6.1 illustrates cell division events that are detected within the sample images. The statistical results about detection accuracy[1] and recall[2] values are listed in Table 6.1. In comparison with the works (Li et al., 2010) and (Huh and Chen, 2011), both of which rely on learning-based techniques for detecting cell-division events, the proposed approach has achieved comparable accuracies.
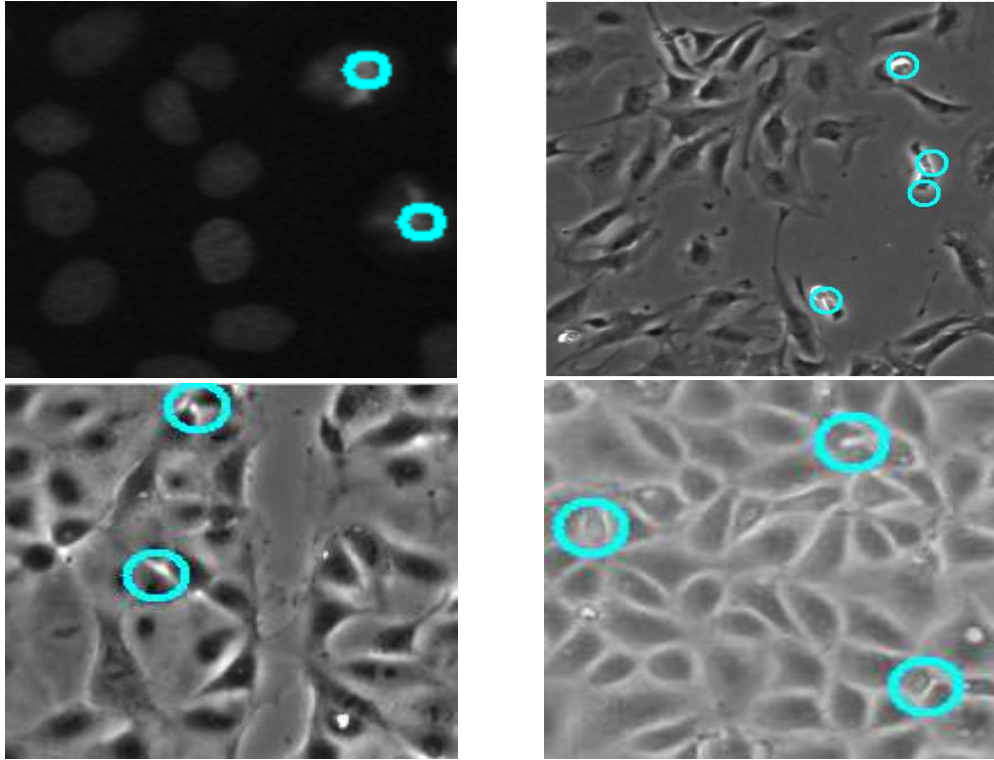
Fig. 6.1: Illustration of cell divisions being detected in the circle-marked regions. From top to bottom, left to right, there are Hela, C2C12, WH and MDCK examples.

---

[1]The number of the correct automatically detected number divided by the total automatically detected number.

[2]The number of the correct automatically detected number divided by the total manually annotated number.

Table 6.1: Quantitative results for cell division detections

	HeLa	C2C12	WH	MDCK
No. of frames	54	200	104	150
Total division events (manually annotated)	69	98	21	51
Correct division detected	58	89	19	42
Non-division detected	10	11	11	24
Recall	84.1%	90.8%	90.5%	82.4%
Accuracy	85.3%	89.0%	63.3%	63.6%

Compared with the high accuracies and recalls that have been achieved within the HeLa and C2C12 image sequences, the proposed approach has reported relatively lower detection accuracies for the WH and MDCK datasets. The main reason is that a considerable number of cell death events (within WH and MDCK) occur and there is much cell debris (in MDCK), and therefore this affects the detection accuracies. Figure 6.2 illustrates that some cell death events and cell debris are wrongly detected as mitosis events. However, the mean recall value of the four tests is 87.0%, which reflects that most of the division events have been successfully identified.
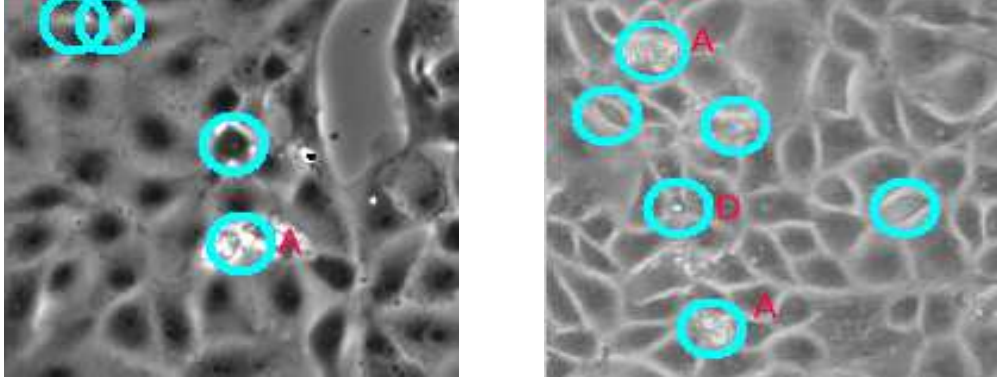
Fig. 6.2: Illustrating that some cell death events and cell debris are wrongly detected as mitosis events in the WH and MDCK images (from left to right). Red "A" letters mark apoptosis events (i.e., cell death), and "D" letters mark cell debris. The circles without red letters beside denote correctly detected mitosis events.

With respect to the performance of the proposed approach under several challenging conditions, there are several clarifications: 1) In the case of a large spatial location change during a cell division process. This situation is reported in both of the HeLa and the C2C12 datasets, which have relatively lower time resolutions than the other datasets. In the HeLa case, newly born cells can have large location deviations from their mother cell's spatial location. The proposed motion estimation method will match the mother cell region into one or both of the children cells (See Figure 6.3). For the C2C12 dataset, large location changes usually take place when a mother cell has not started splitting itself. This case will be detected as a large motion, and therefore will not affect the robustness of the proposed approach for cell division detection. 2) OF based motion estimation can be tolerant to a small ratio of a cell region appearing/disappearing (due to cell growing/shrinking). So, our approach works better in detecting occlusions that are caused by a relatively larger ratio of a cell region being occluded. By observation, it has been found that division cells indeed have much larger ratios of region occlusion than growing/shrinking cells. So, the motion occlusion can be effectively detected by the proposed method. 3) Distinguishing cell overlapping and cell division events is currently out of the scope. Furthermore, because the image sequences being tested are taken from in-vitro cells that are cultured within considerably thin

containers, so the rate of cells that heavily overlap with each other is small. 4) The proposed approach is robust in the presence of scattered noise, due to motion regularisation that is embedded within the OF estimation algorithm.
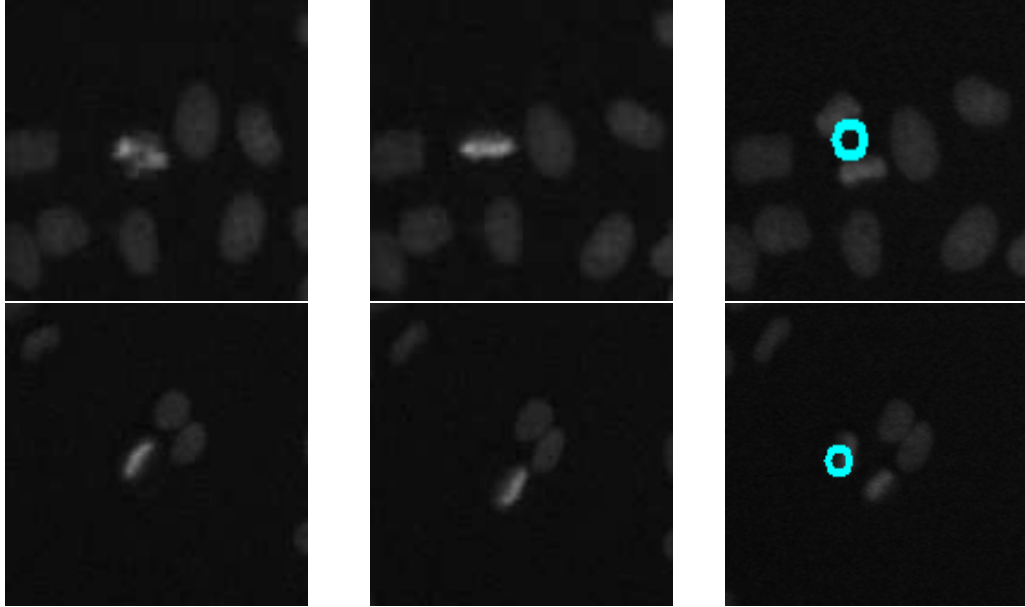
Fig. 6.3: Illustrating two (Hela) mitosis events. In the top row, the mother cell region is matched onto the centroid region of the two children cells. The bottom row shows that the mother cell is matched onto one of the children cells.

To summarise, the generalised approach proposed in this chapter is achieving good performance levels, especially when it is considered that the approach is being tested under challenging conditions, where the cell division patterns are very different in each of the four image sequences.

6.4 Conclusions

In this chapter, a novel cell-division detection approach has been described. The proposed approach adopts a motion-based occlusion detection strategy, which is quite different from existing segmentation or training based methods. The major contribution of this work is that the proposed approach can success-fully detect dividing cells with a variety of division behaviour. This has been

proven by our experimentation on four different types of image sequences.

Chapter 7

A Model and Motion based Dense Cell Population Tracking Framework

Intense current research requires quantitative analysis of cell behaviour in dense cell populations. The low contrast cellular image quality, diversity of cell shapes, frequent cell interactions, and complex cell motion all pose significant problems to the efficient and robust cell tracking in phase-contrast cellular images. Motivated by this, the chapter proposes a novel automated cell tracking framework, where two main tasks are targeted: the segmentation of cell shapes in low contrast phase-contrast images, and the estimation of cell trajectories in dense and highly dense populations. Specifically, the parametric active contour model (also called snake model) and the optic-flow (OF) technique have been seamlessly combined to cooperate with each other, so as to address the shortcomings of each, in the process of simultaneously tracking cell deformations and movements. Experimental results have proven that the proposed approach is able to achieve accurate and robust results, in dealing with several particular challenges, including the accurate segmentation of cells with ambiguous boundaries, tracking large displacement cells, accurately segmenting and tracking partially overlapping cells, and the consistent tracking of elongated cells. Moreover, the cell tracking performance of the proposed approach is superior to the state-of-the-art motion based approach. The pro-

posed method has also achieved comparable results in comparison with the leading-edge model based cell tracking approaches, which the proposed approach involves neither training processes nor heavy post-processing steps.

7.1 Introduction

7.1.1 Problem Summaries and Challenges

The fundamental purpose of this chapter is to track dense cell populations in phase-contrast image sequences. Regarding this task, there are four subproblems to be addressed.

Because of the specific illumination technique that is employed in phase-contrast microscopies, cells usually appear as relatively dark regions surrounded by bright halos, which lead to flexible boundary candidates. Figure 7.1 shows multiple boundary candidates (that are manually marked by different experts) for individual cells in image samples. Also note that the gray level of the image foreground (i.e. cell regions) can be very similar to that of the background region in phase-contrast images. Due to the low contrast between cells and the background, some cells seem to have "broken" or blurred boundaries. These issues pose great difficulties in accurately determining or outlining a whole cell region in phase-contrast datasets.
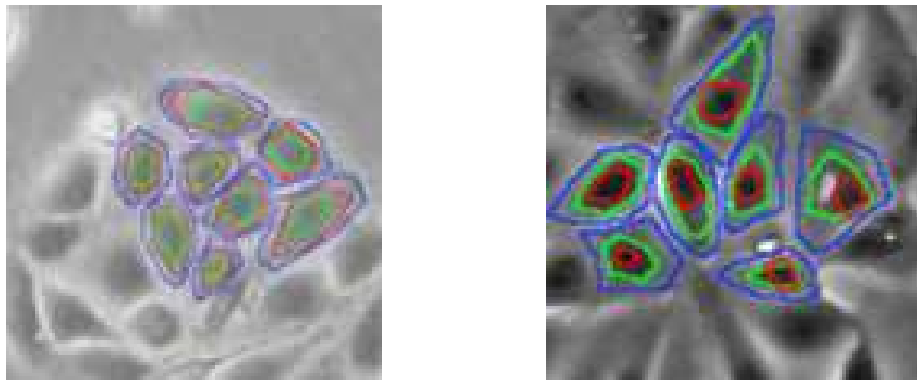
Fig. 7.1: Illustrating ambiguous boundaries of cells in phase-contrast image samples, with the three different colours of contours being manually outlined.

The problem of tracking cells with large displacement and/or deformation will be investigated. In order to make cells have small displacements between consecutive frames, the time resolution of image sequences needs to be set sufficiently high, i.e. a high frame rate of image recording is needed. However, a high sampling rate of images will affect cells' lifespan and behaviour. This is due to the fact that, by capturing images of cells, it will damage the cells. So, the true development of the cells can not be obtained. That is why low time-resolution image sequences are particularly considered in this work. The cells will thus appear to have large displacement and deformation. This, however, poses a large level of complexity in tracking.

Since the tracking targets are highly concentrated cell populations, an immediate difficulty is how to detect or segment cells that are in a close neighbourhood, and cells that partially overlap with their neighbours. It is also non-trivial to obtain accurate cell association results between adjacent frames.

In addition, another challenging problem is the tracking of elongated, long shape cells (see illustrations in Figure 5.6 in Section 5.3.3). In different types of cellular datasets, or even in the same dataset with a dense population, cells may have a large variety of shapes. Compared with relatively round shaped cells, elongated cells are more difficult to be accurately segmented or consistently tracked. In existing cell tracking works, a considerable rate of tracking lost results from elongated cells. This is because a higher rate of over-segmentation happens on elongated cells. Even worse, the segmentation and tracking error of this type of cells can be easily accumulated over a number of frames, and thus leads to lost tracking.

7.1.2 Contribution Summaries

The main contribution of this chapter is a novel model and motion based cell tracking framework. The complementary properties from two different categories of cell tracking approaches have been taken advantage of, and the shortcomings of each side are suppressed. Thanks to the proposed approach, many challenging problems have been accommodated or solved, including accurately segmenting and tracking partially overlapped cells, consistently tracking

elongated cells, and the tracking of large displacement cells (within low time-resolution image sequences). It is worth noting that these problems have not been well solved or are usually ignored in existing works, due to the common cell segmentation and tracking difficulties under the conditions of dense cell populations and phase-contrast image sequences.

The second contribution is that the dynamic directional, gradient vector flow (DDGVF) technique is, for the first time, introduced into the task of segmenting and tracking cell populations in phase-contrast datasets. The application of DDGVF based snakes plays an important role in precisely outlining ambiguous cell boundaries within phase-contrast images, and correctly segmenting and tracking cells in clustered environments.

7.1.3 Organisation

The chapter is organised as follows. Section 2 declares the motivation of combining model based and motion based cell tracking approaches. Then, some most related works are briefly reviewed. Section 3 compares important, candidate algorithms/techniques that are considered in the construction of the proposed approach. In Section 4, the overall structure and the main function blocks of the proposed system are explained. After that, the mathematical formulation is provided for the developed core algorithm that underpins the proposed framework. Section 5 includes the experimental results that validate the the proposed approach on real cellular datasets.

7.2 Why a Model and Motion based Hybrid Approach

Recall that in Chapter 5, cell tracking approaches have been categorised into three groups: segmentation-association based, model based, and motion based approaches. In this section, the most related cell tracking works are briefly reviewed. Particularly, further focused discussions are present around the following two topics.

1. The motivation of choosing the model based cell segmentation technique.

According to the survey in Chapter 5, accurate cell detection/segmentation results are vital for the success of all of the three categorised approaches. For this reason, it is necessary to compare the performances of cell tracking approaches that are driven by low-level and high-level segmentation algorithms, especially in the tracking of dense cell populations within phase-contrast datasets. Methods such as thresholding and water-shed (Vincent and Soille, 1991) rely on low-level image information for segmenting cells, without taking into account the basic properties of cell shapes (e.g., edge continuity and smoothness). Although those methods can usually achieve good performances in segmenting fluorescence cellular images, they suffer from a considerable rate of under- or over-segmentation in the case of dealing with phase-contrast datasets (see works of Kaakinen et al. (2014), Dewan et al. (2011) and Yang et al. (2006)). By contrast, model based cell segmentation algorithms, such as parametric or non-parametric active contours (Kass et al., 1988, Caselles et al., 1997, Kichenassamy et al., 1995), naturally allow the encoding of shape prior constraints about cells (Ray and Acton, 2004, Padfield et al., 2009, Jurrus et al., 2009). So, model based methods can potentially produce better estimates of cell morphologies (Li et al., 2008), and are generally not constrained by particular types of image datasets. Apart from that, extra high-level priors, such as the repulsive constraint, can be intrinsically integrated into the cell model, modelling the repulsive relationship between cell membranes. Those constraints have been proven effective in preventing incorrectly merged segmentation (see (Zimmer et al., 2002, Li et al., 2008) for demonstrations). Therefore, model based approaches are relatively more suitable for tracking dense cell populations within phase-contrast image sequences.

2. The complementarities between model and motion based cell tracking approaches.

Model based approaches, such as parametric or non-parametric active contours, are preferred to be applied to tracking small displacement cells. Because of the local minimisation problem, a model based approach can not track a cell that has insufficient or no overlap with its translocated cell, between two consecutive frames. Due to the common difficulty in tracking large

displacement cells, in most of the existing works, the time resolutions of the tested image sequences are set considerably high. By doing so, the problem of tracking large displacement cells is side stepped. However, a long time or a high frequency of image recording will affect cells' lifespan and behaviour. By contrast, motion based approaches (A.J. Hand and MacNeil, 2009, Yang et al., 2008), if being embedded into a pyramid or multi-scale implementation framework, can be easily utilised to track large displacement cells. So, motion estimation techniques, such as OF algorithms, are good candidates to assist model based approaches in tracking large displacement cells.

The problem of segmenting and tracking partially overlapped cells is not well solved in the literature. In the case of two (or more) cells partially overlapping with each other, parts of the boundaries will be missing. The repulsive mechanism, when embedded in model-based cell segmentation and tracking approaches, is useful to temporarily prevent incorrectly merged segmentation. However, the cells' interface can not be accurately located by only using model based approaches, which rely on the spatial-space information for segmentation. Again, the utilisation of the motion information (i.e. the temporal-space information), derived from cell boundary and intercellular regions, is reasonable to improve the accuracy of cell region and centroid measurements for partially overlapped cells.

Thirdly, elongated, long shape cells are commonly seen in different cellular datasets, while this type of cells are relatively difficult to be tracked consistently. Region based cell models (e.g. parametric and non-parametric active contours), because of suffering from the leaking problem, might not be suitable for the task of tracking cells that have blurred boundaries and are in dense populations. Cell membrane models are usually based on parametric active contours. Due to the reliance on boundary based feature matching, parametric active contours have the issue of ambiguous feature association, especially for tracking elongated cells that have relatively low curvature boundaries. Given this problem, OF algorithms or image registration methods can be exploited to provide complementary feature-matching information from intercellular regions, so as to address the model based tracking process.

Conversely, from the side of motion based approaches, in order to address

the error accumulation issue, motion based methods need frequently refined cell segmentation and tracking results, which can be provided by model based approaches (A.J. Hand and MacNeil, 2009).

One other problem is that, for the purpose of recovering (relatively) large motion, the image down-sampling rate, in the multi-scale motion estimation framework, should be sufficiently large. This process might degrade detailed cellular motion. However, model based cell segmentation and tracking approaches are good at capturing small motion of cell boundaries.

Therefore, model based and motion based cell tracking approaches are complementary to each other in many aspects. It is thus very promising to construct a powerful new cell tracking approach, by collecting the merits from both worlds.

7.3 Comparison of Candidate Algorithms or Techniques

7.3.1 Parametric Vs. Non-Parametric Active Contours

Active contour models (ACMs) can be divided into non-parametric and parametric forms. Parametric active contours, or snakes, are represented by curves with explicit descriptors, i.e. closed or open curves (connected by a number of control points). It is a form of single object model, thus cannot inherently deal with topological biological changes (such as splitting and merging). Non-parametric active contours can be defined implicitly, such as the zero-th level-set of an evolving curve, that does not require explicit parameterisations. Thus, non-parametric active contours do not suffer from any constraints on the topology (Zimmer et al., 2002). Comparing these two types of ACMs, the following aspects are noticed:

- Non-parametric active contours can more accurately detect arbitrary object shapes because of the parameter-free representation. However, in the case of individual cell regions with non-uniform intensity, and the situation where there are a diversity of cell shapes and sizes, non-parametric

113

active contours can over-segment individual cell regions as fragments. Also, they are generally more sensitive to image noise and local minima when the image contrast is low (Farzinfar et al., 2008). This usually results in unstable cell shapes between subsequent frames and inaccurate boundary segmentations. The reason is that the non-parametric representation of active contours lacks the necessary shape-continuity constraint. In comparison, shape and size constraints can be more easily integrated into parametric active contours.

- One desirable property in the task of tracking highly dense cell populations is the modelling of inter-cellular relationships, such as the repulsion between cells. As single-object models, snakes are relatively more easy to be integrated with the repulsion constraint.

For the reasons as discussed above, parametric active contours are chosen as the major tool for modelling cell shapes (or membranes) in the current work.

7.3.2 Different External Forces based Snakes

Recall the introduction in Section 2.2.1 that a standard snake model usually contains an external energy term, which has the effect of driving snakes toward desirable features, such as bright/dark image regions, or large gradient regions (i.e. object edges). So, the external energy term essentially encodes the appearance property of an object that is to be segmented or tracked. Thus for designing a particular ACM model, external energy terms play important roles. Within the literature of external forces or energies of snakes, the region based (Chesnaud et al., 1999, Xie and Mirmehdi, 2004) and the boundary based (Kass et al., 1988, Ray et al., 2002, Xu and Prince, 1998b) are two main groups.

In region based ACM models, the inner and the outer regions that are defined by snakes are both considered (Chesnaud et al., 1999). Thus, they are relatively more robust to image noise and can be well adapted to segment or track cells with interior textures. However, in a dense cellular environment with phase-contrast images, the most outstanding challenges, in the segmen-

tation and tracking task, are due to the clustered cells, blurred cell boundaries, as well as the similar gray-level distributions between the foreground and the background image regions. It is thus difficult to effectively model the statistics of cell regions. For this reason, region based snakes suffer from a serious leaking problem. The shape and size constraints, as exploited in the works of Ray et al. (2002), Padfield et al. (2009), might be considered to relieve the leaking problem. However, the existence of a large variety of cell shapes (in dense populations will make these constraints less effective.

Among boundary based snake forces, the gradient vector flow (GVF) based force (Xu and Prince, 1998b) is popular, due to the fact that GVF largely increases the capture region of snakes. Ray and Acton (2004) propose a motion gradient vector flow (MGVF) that modifies the GVF technique by taking into account the moving direction prior of a target. MGVF based snakes can thus be applied to tracking relatively large displacement objects. Some other examples of external forces are proposed by Cohen (1991), Wang et al. (2009), Gil and Radeva (2003), Li and Acton (2007), etc. Those external forces based snakes have been successfully used in tracking individual targets, or a non-dense population. It is interesting to note that, in (Cheng and Foo, 2006), a so-called, dynamic directional gradient vector flow (DDGVF) is developed. DDGVF makes use of both positive- and negative-step edges (i.e., edges of positive/negative gradients along x or y axis) inside an image, and computes separate force fields with respect to the four types of edge maps. DDGVF thus takes advantage of both gradient direction as well as gradient magnitude information of image edges. This is unlike other external forces, that only consider gradient magnitudes, in the construction of external force fields. By this means, DDGVF snakes can potentially lead to reliable segmentation results where clustered edges exist.

Taking an overall consideration, it is determined to employ a boundary based snake model, with a DDGVF based external force.

7.4 Proposed Cell Tracking Framework

7.4.1 Main Function Blocks

This chapter implements an automated cell segmentation and tracking system, which consists of the following major components: image preprocessing, low-level image segmentation, snake initialisation, snake based cell tracker, motion based cell tracker, DDGVF based external-force map generator, and measurements output.

1. Image pre-processing: Currently all that is used at this stage is a median filter to slightly smooth each input image, removing impulsive noise without blurring edges.

2. Low-level image segmentation: An extended minima transform (Soille, 2003) operation is exploited to roughly separate cells from the background. The resulting binary mask, as shown in Figure 7.2(a) contains a number of white blobs that represent connected cell regions of pixels with similar intensities. The chosen segmentation method is effective because the cells that are recorded by the phase-contrast microscopy appear as dark regions surrounded by bright halos. In order to handle the problem of pre-segmenting cells with a different appearance, other methods can be also employed. According to practical experiments, a so-called divergence-free GVF based approach is also a good choice to fulfill the task of image pre-segmentation.

3. Snake initialisation: Multiple snakes are automatically initialised in the first frame as small green circles (Figure 7.2(b)) that are located located at the centroids of the bright blobs in the binary map. In practice, the average radius of the cells is empirically set in this initialisation process.

4. Snake based cell tracker: Cell regions are segmented by snake based trackers in this block. With exception of the first frame, cell boundary contours according to the segmentation and tracking results in each previous frame are used as the new initialised trackers to automatically track target cells across subsequent frames.

5. Motion based cell tracker: For every two adjacent frames within the image sequence, a forward-backward OF estimation process is applied in this block. The calculated OF fields serve two purposes:

- Motion based cell tracking: Based on the tracked cell regions in the last frame, the forward OF field maps the regions into the current frame, resulting in predicted new positions and regions for the cells. The predicted information is then exploited for the cells as reference boundary/outline positions, which are fed into the snake based cell trackers. In essence, the reference boundary positions work as a constituent part of the designed, model&model coherence-constraint based, external energy term (for the snake model). More explanations related to the new energy term will be provided in the core algorithm section.

- OF confidence estimation: A mutual consistency checking process (Alvarez et al., 2007, Brox et al., 2004) is conducted between the forward and backward OF fields. This outputs an OF estimation confidence map, the role of which is the other constituent part for the coherence-constraint based external energy term. To be more concrete, this confidence map provides the weighting parameter for the coherence-constraint energy term.

6. DDGVF based force map generator: This block is responsible for calculating boundary-based external forces for snake based cell trackers.

7. Measurements output: This includes the information of the estimated cell regions and cell centroids for in each image involved.

7.4.2   Core Algorithm

At the heart of the proposed core algorithm is a parametric active contour model, that contains two traditional internal energy terms, and three external energy terms. The snake is represented by a close curve $C(s) = (x(s), y(s))$, which is parameterised by s in the range of $[0, 1]$. In practice, a number of discrete control points (called snaxels) shape the curve. The snake works by

117
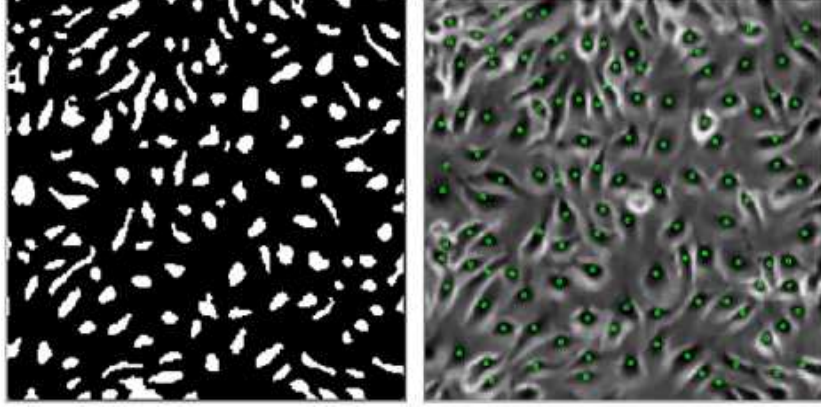
Fig. 7.2: Left: the binary map from image pre-segmentation; right: automated snake initialisation.

being nested inside the scheme of minimising the energy functional $E(C)$ as formulated below:

$$E(C) = \int_1^0 (\alpha |C_s|^2 + \beta |C_{ss}|^2 + E_{1,2,3}^{ext}(C(s)))ds \tag{7.1}$$

where the snake internal energy, consisting of the first two terms, is defined as the same as that in the traditional snake model of Kass et al. (1988). $E_{1,2,3}^{rep}$ represents the bonded set of the designed external energies. As mentioned in Section 2.2.1, the snake that minimizes $E(C)$ must satisfy the Euler equation: $\alpha C_{ss} + \beta C_{ssss} + \nabla E_{1,2,3}^{ext} = 0$. These three terms are viewed in order as the elastic force F_{ela}, bending force F_{cur}, and the bonded external force F_{ext} applied on the snake. The snake will move because of the competition between the forces, and will reach equilibrium when the forces are balanced by each other. In this report, the novelty of the proposed snake relies on the design of $E_{1,2,3}^{ext}$, which contains three sub-components:

$$F_{ext} = \omega_1 F_{ddgvf} + \omega_2 F_{coh} + \omega_3 F_{rep} \tag{7.2}$$

where F_{ddgvf} is the boundary based force, F_{coh} stands for the model&motion coherence force, and F_{rep} denotes the repulsive force between snakes that models cell-cell repulsion. $\omega_1, \omega_2, \omega_3$ are weighting coefficients for the corresponding external forces. To be more concrete, the three forces are described as below:

1. The boundary force F_{ddgvf} is responsible for driving snakes towards cell boundaries. It is derived from DDGVF (Cheng and Foo, 2006). For

118

each image involved, there are four DDGVF fields generated, which are respectively calculated in terms of two positive step edges (along the positive directions of x and y axis) and two negative step edges (along the negative x and y directions). DDGVF thus exploits the directional information of image gradients as well as gradient magnitudes, for the construction of the boundary based force fields. DDGVF can be well applied in situations where clustered edges exist. See (Cheng and Foo, 2006) for detailed implementation steps.

2. The model&motion coherence force F_{coh}, defined as $F_{coh} = \int_1^0 (C(s) - \tilde{C}(s))ds$, with C representing the current snake, $\tilde{C}$ standing for the predicted reference boundary for the snake. The predicted reference boundary is provided by the motion based cell tracker. So, this force encourages the tracking results of snakes and OF based motion estimators to agree with each other. Moreover, the OF confidence map (calculated also in the function block of the motion based tracker) determines the weighting parameter ω_2 for this force F_{coh}. By this means, in regions where OF confidence is high, the coherence between the snake and the motion based trackers is strongly enforced; otherwise, a weak coherence constraint is imposed. F_{coh} and ω_2 thus constitute the designed model&motion coherence term. The implementation method for motion estimation and the steps of mutual consistence checking can be referred to Section 6.2.1.

3. In dense cell populations, since cells are commonly seen clustered together, so nearby snakes may converge onto wrong targets. The repulsive force F_{rep} is thus introduced for each snake to deal with this situation, by inverting the image force in the regions that are occupied by the neighbouring snakes. The idea is inspired by the work of Zimmer et al. (2002).

7.5 System Evaluation

7.5.1 Cellular Datasets

To demonstrate the robustness and validity of the proposed approach, the system is applied to Bovine Aortic Endothelial Cells (BAEC) and Madin Darby Canine Kidney Epithelial cells (MDCK) in phase-contrast image sequences, with their spatial resolutions respectively 800×600, and 688×512. A representative cropped (with a resolution of 400×300) and magnified BAEC image is shown in Figure 7.3, particularly for illustrating the relatively poor image quality, weak cell boundaries, and highly dense cell population.
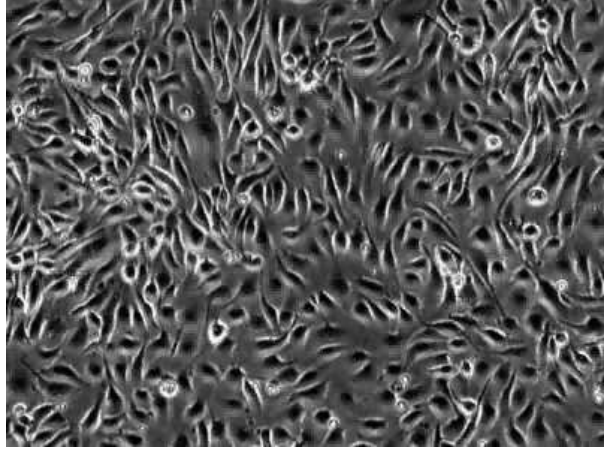
Fig. 7.3: Phase-contrast BAEC image — a cropped and magnified version.

It should be noted that, although many cell segmentation and tracking works have been reviewed in Section 5.3, cellular datasets in those works are usually not publicly available. This has posed a difficulty in performance comparison. However, in this work, more challenging cellular datasets have been selected for the system evaluation purpose. Specifically, higher densities of cell populations are considered, so the targeted images are relatively more difficult to segment. Apart from that, the time resolutions of the targeted image sequences are usually 3 or 4 times lower than that of the related works. So, cells usually have relatively large displacements and deformations between consecutive frames.

The following subsections present the segmentation and tracking results that are returned by the proposed cell tracking system.

120

7.5.2 Qualitative and Quantitative Segmentation and Tracking Results

In dense cell populations, frequent cell-cell interactions can easily cause incorrect segmentations, when adjacent snakes merge together. For this reason, the repulsive mechanism is embedded into the snake model. Figure 7.4 demonstrates how two cells (the 46th and the 47th cells), that have very close contact over several frames (from the 3rd frame to the 10th frame), are separately segmented and tracked.
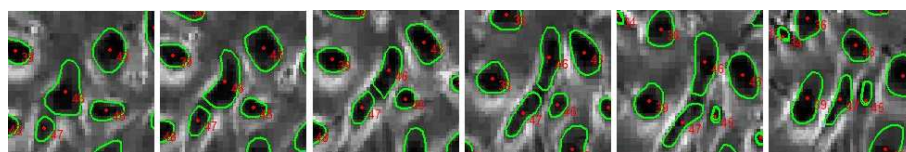
Fig. 7.4: The six cropped images show that two interacting (i.e., partially overlapping) cells (the 46th and the 47th cells) are successfully segmented and tracked over a number of frames.

In Figure 7.5, two frames (the 2nd and the 9th frames) show the cell deformation tracking for a BAEC cell population. Note that many elongated-shape cells are tracked over from the 2nd frame to the 9th frame.
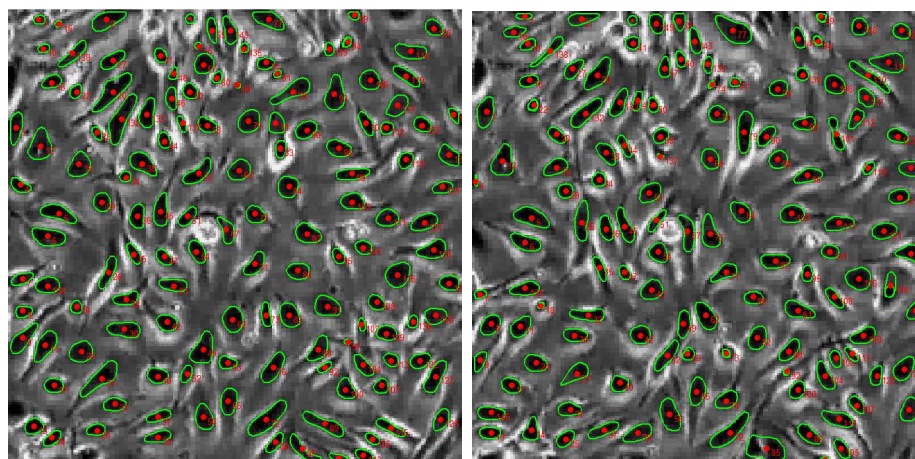
Fig. 7.5: Illustrating the segmentation and tracking process for a BAEC cell population from the 2nd and the 9th frames.

The segmentation results of approximately 50 BAEC cells, and 30 MDCK cells

are shown in Figure 7.6. The illustrations show that the snakes can efficiently segment dense cell populations. Moreover, it can be seen that the designed snakes have accurately segmented ambiguous, and halo-region-surrounded cell boundaries. The result is thus consistent with the expectation that DDGVF snakes are powerful in segmenting cells with flexible boundaries and in clustered environment.
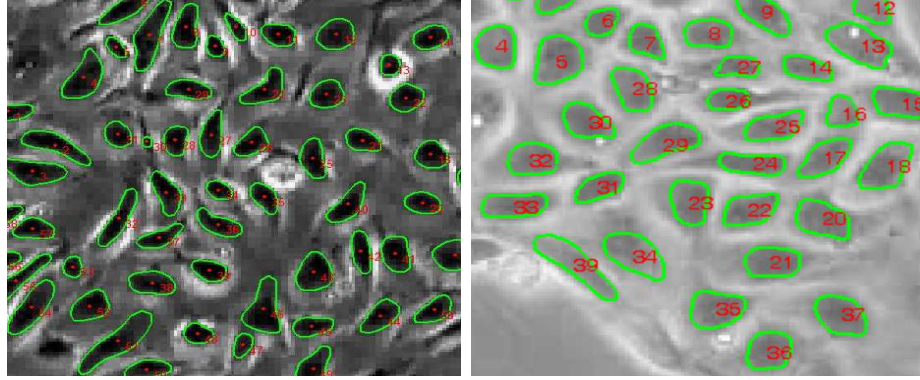
Fig. 7.6: The segmentation results of about 50 BAEC cells (left), and 30 MDCK cells.

Due to the currently unavailable ground-truth data with quantitatively segmented cell regions, Figure 7.7 presents a segmentation result for a MDCK image that is obtained by the h-maxima transform based algorithm, which is a representative tool in cell segmentation-association based approaches (see (Dewan et al., 2011, Thirusittampalam et al., 2013)). The h-maxima transform approach has been acclaimed very efficient in detecting cells in phase-contrast image sequences. It is worth emphasising that the segmentation result, displayed in the left image in Figure 7.7, needs a number of tuning to obtain optimised values, by using the h-maxima transform method. According to a qualitative comparison between the left and the right image in Figure 7.7, it is concluded that the proposed DDGVF snakes can achieve superior cell segmentation results.
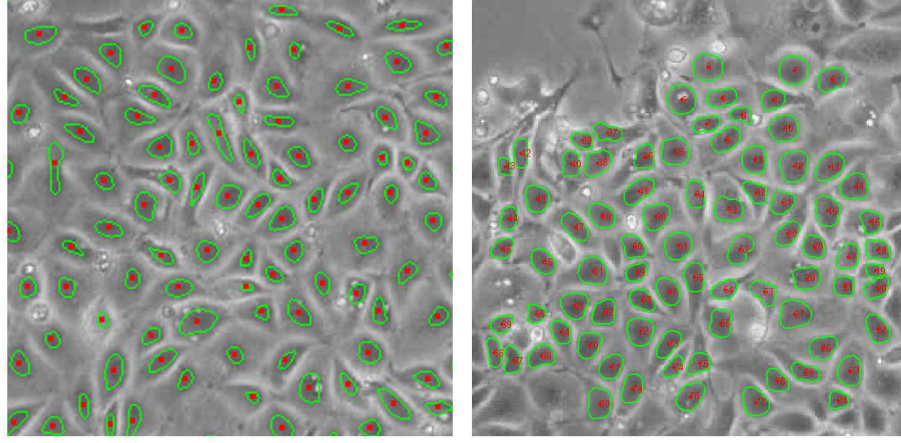
Fig. 7.7: Both are MDCK cellular images. The cells in the left image are segmented using the h-maxima transformation. The right image is segmented by the DDGVF active contours.

7.5.3 Overall Cell Tracking Accuracies

To numerically quantify the accuracy of the proposed cell tracking approach, the automated tracking results are compared against the manually annotated data that is established by an expert. The tracking accuracy, used in this section, is given by the number of correctly tracked cells that are identified by the proposed approach with respect to the total number of cells that have been taken into account in the first image.

Note that, in Equations (7.1) and (7.2), there are five important coefficients. α and β, controlling the weights of the shape-related energies of the ACM, are intuitively assigned. In order to facilitate the coefficients tuning for the external forces, F_{ddgvf}, F_{coh} and F_{rep} have been normalised. In practice, very limited tunings are usually needed for the five coefficients, and a typical set of values for these five parameters has been shared for tracking different types of cellular datasets. Usually, very limited tunings are needed for the five coefficients.

Table 7.1 lists the cell tracking accuracies for both of the BAEC and the MDCK dense cellular datasets.

Table 7.1: Quantitative results for cell tracking accuracies

	No. of frames	No. of cells	No. of tracked cells	Accuracies
BAEC	50	131	112	85.5%
MDCK	56	79	69	87.3%

Figure 7.8 shows six BAEC image frames (frame No. 11, 13, 15, 17, 19, and 21) out of the whole 50 frames, illustrating the cell tracking process over the image sequence. Among the 19 lost cell-tracking cases, there are 3 cells failed to be tracked due to segmentation errors, 3 cells due to occlusions, 3 cells because of mitosis events, 5 lost cases caused by elongated cells, and 5 cells due to large deformation and abrupt displacements. Because of the low time resolution of this image sequence, cells usually have large displacement and/or large deformation. There is also a large rate of elongated cells in the BAEC image sequence. Cell segmentation and tracking in BAEC datasets is thus very challenging, and existing works usually avoid using this type of datasets. However, the proposed approach has achieved a high overall tracking accuracy 85.5%. Specifically, according to the experimental results, more than 60% large deformation/displacement cells (5 out of 8), and more than 50% elongated cells (5 out of 9), have been successfully tracked over the whole image sequence.

The MDCK dataset contains a highly dense cell population. So, the foreground and the background image regions are non-trivial to be segmented by purely low-level image processing tools. Figure 7.9 illustrates the tracking process for the MDCK cells, with the six frames (frame No. 10, 16, 21, 28, 33, and 38), displayed. There are 79 cells initialised, and the proposed approach has successfully tracked 69 cells over the whole image sequence. Among the 10 lost tracking cases, 8 cases are caused by cell division, 1 lost tracking because of temporal occlusion, and 1 case due to large deformation. Since dividing cells in the MDCK dataset usually have large intensity changes, a high rate of lost tracking results from cell division events.
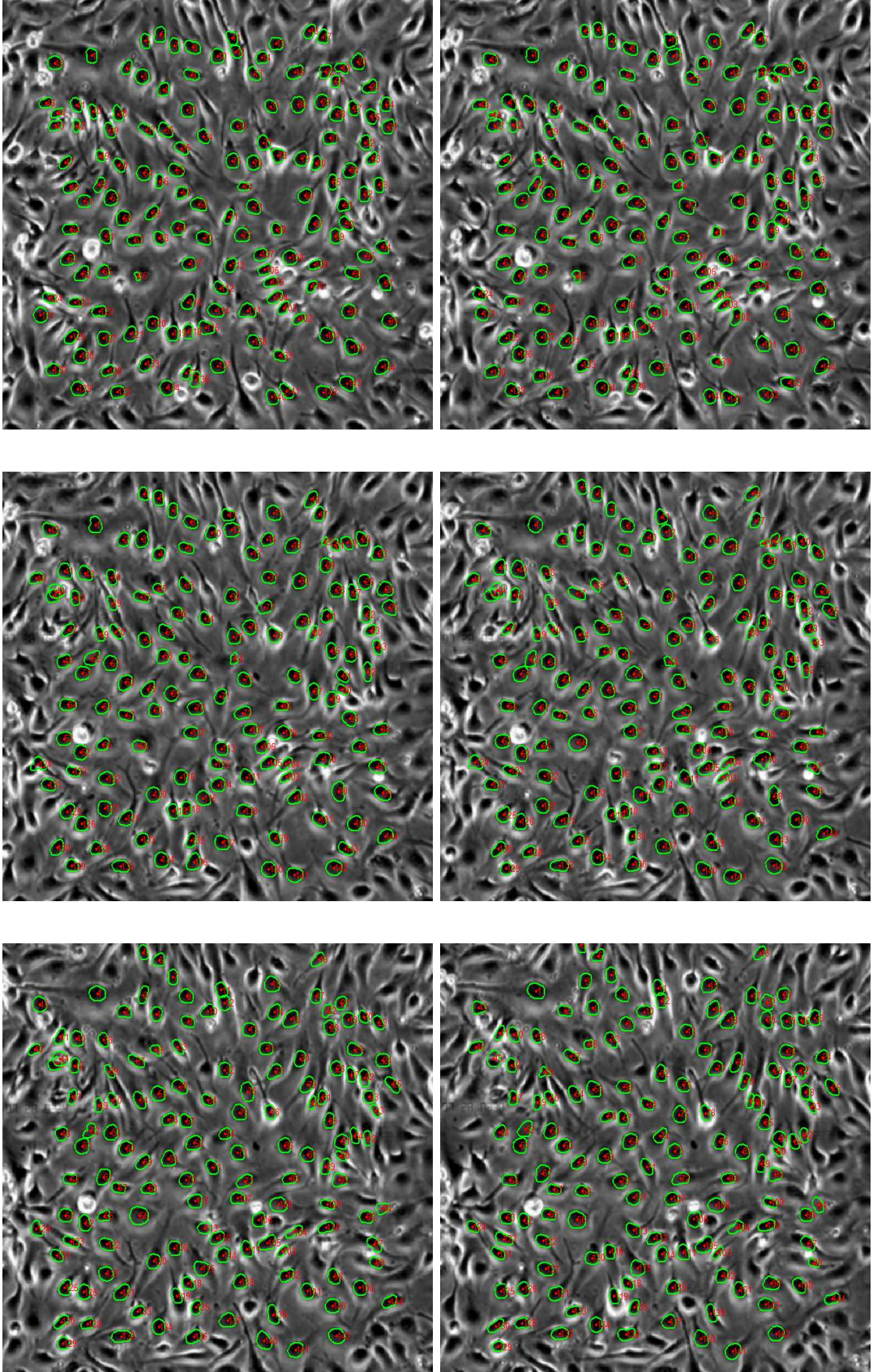
Fig. 7.8: From top to bottom, left to right, frames No. 11, 13, 15, 17, 19, and
21 illustrate the cell tracking process for the BAEC cell population.
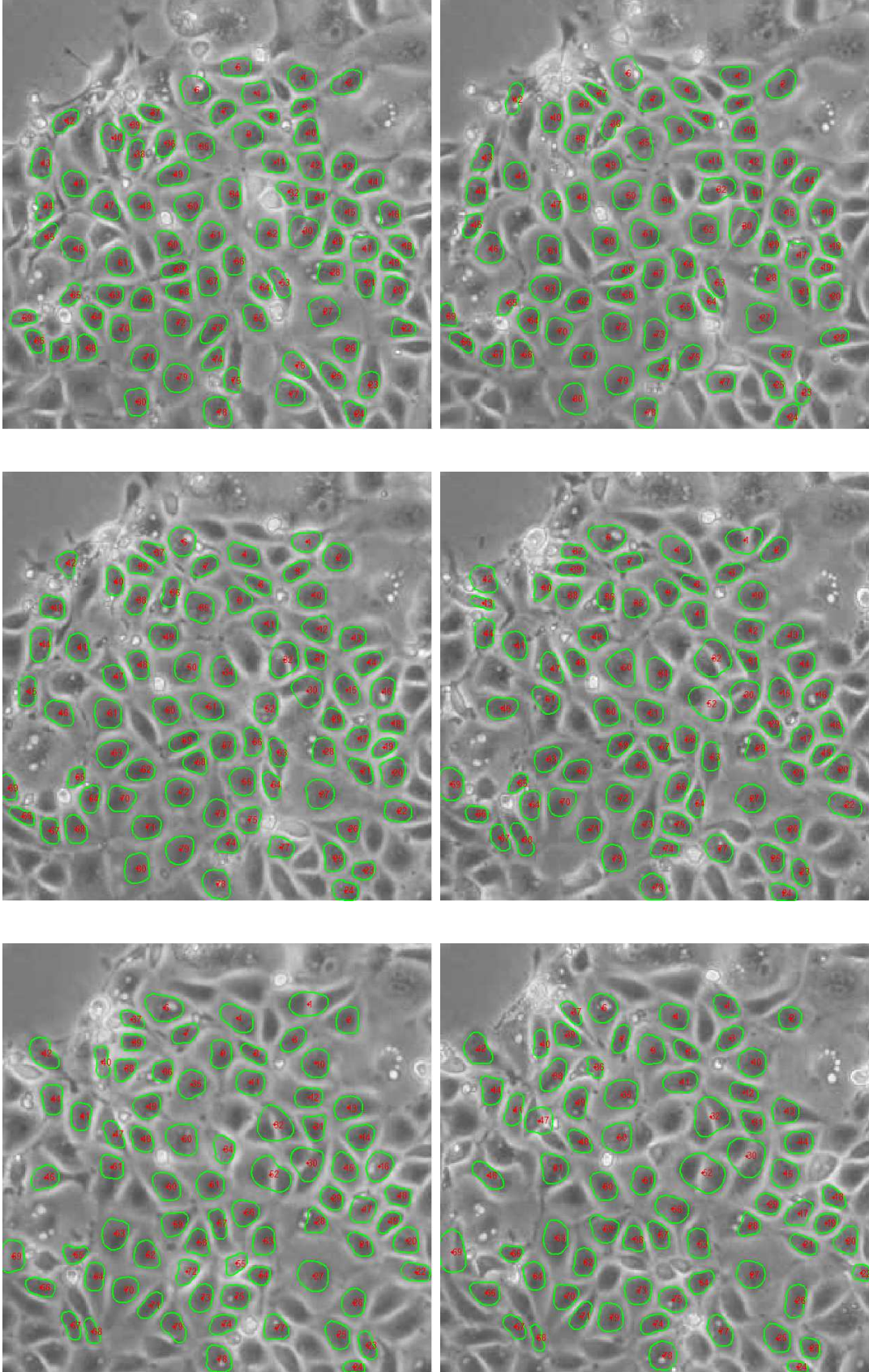
Fig. 7.9: From top to bottom, left to right, frames No. 10, 16, 21, 28, 33, and 38 illustrate the cell tracking process for the MDCK cell population.

In the testing result for the MDCK dataset, 8 out of 11 dividing cells fail to be tracked. The tracking accuracy including cell division events is 87.3% (as listed in Table 7.1). When ignoring division events, the tracking accuracy has achieved 97.1% (i.e., 66 cells out of 68 being tracked). These two types of tracking accuracies are consistency with the existing literature, where some works only consider tracking accuracies of non-division cells and some others take cell division events into account. So, the proposed approach has been proved very stable in segmenting and tracking cells in a highly dense environment.

The tracked cell-centroid positions are recorded along the image sequences. This has led to 2-D and 3-D cell trajectory maps. For an intuitive illustration, Figure 7.10 displays the 2-D and 3-D trajectories for about 130 BAEC cells being tracked over 50 image frames.
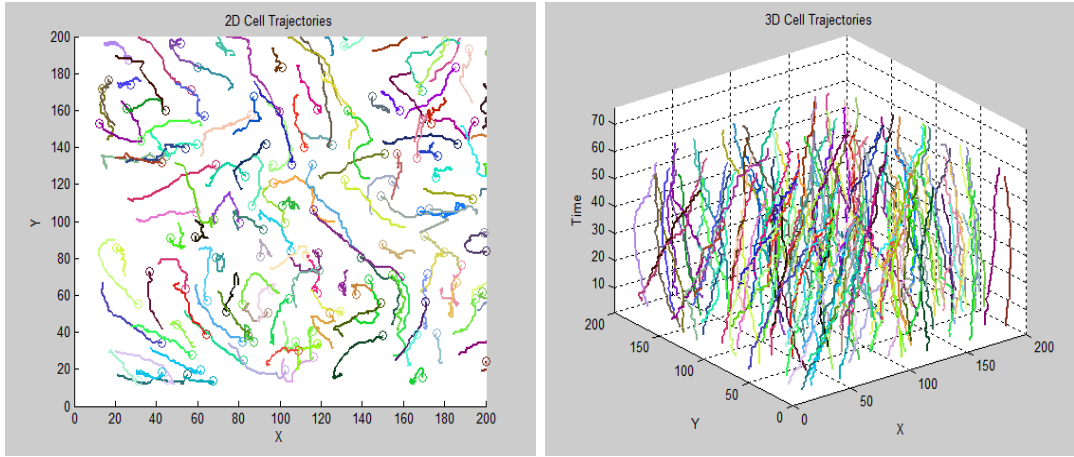
Fig. 7.10: Illustrating the 2-D and 3-D trajectory maps for a BAEC cell population.

Comparable with the most related, state-of-the-art cell tracking approaches, the proposed method has achieved very impressive performances. To be more concrete,

- In comparison with the model-based cell tracking approach of Li et al. (2008), the proposed approach has achieved comparable results. The reported tracking accuracies of (Li et al., 2008) are in the range of $86.9 - 92.5\%$. However, it should be noted that their approach involves a sophisticated, trained motion-filtering for cell tracking, and a post-

processing strategy for re-connecting broken cell trajectories. In contrast, the proposed approach does not rely on any training processes nor extra post-processing steps.

- In the motion-based cell tracking approach of A.J. Hand and MacNeil (2009), much sparser cellular datasets are used, and their tracking method works well on image sequences with only a small number of frames. So, when compared with the method of A.J. Hand and MacNeil (2009), superior tracking results are enabled in the proposed approach.

- Also, compared with other related cell segmentation and tracking works, such as (Dewan et al., 2011, Chatterjee et al., 2013, Kaakinen et al., 2014), the cellular datasets in this work contain much higher densities of populations. Apart from that, the time resolutions of the targeted image sequences are more than 3 times lower than the related works. These factors have posed large difficulties in the segmenting and tracking tasks. While, the proposed approach has achieved superior performance.

Since the proposed approach has been tested against image sequences with much lower time resolutions than related works, less images need to be captured, and cell damaging can be reduced. This is thus a very important merit for the proposed approach to be applied in practical applications.

7.6 Conclusions

In this chapter, a model and motion based novel cell tracking framework has been developed. The snake based and the motion based cell trackers are seamlessly combined by the proposed model-motion coherence constraint. The DDGVF based technique has been, for the first time, applied to the tracking of dense cell populations over phase-contrast image sequences. After that, qualitative as well as quantitative cell segmentation and tracking results have been reported, by testing the proposed approach on challenging real cellular datasets. The experimental results have confirmed that the proposed approach can achieve superior performances, in comparison with the state-of-the-art approaches.

Chapter 8

A Sparsity&Non-Sparsity Constraints based Prior-Adaptive Regularisation Approach

Optimal motion-distribution priors (MDPs) are of crucial importance for accurate optic-flow (OF) estimation. Existing OF regularisation schemes usually empirically choose a globally fixed MDP over the whole motion field. Common choices of MDPs, in OF regularisation approaches, are TV based sparse priors, or Tikhonov based non-sparse priors. Globally set MDPs might be helpful to enforce proper global statistics in the considered motion space. Global MDPs, however, usually either result in a piecewise blocky motion field (the so-called staircase effects), or conversely they over-smooth motion boundaries or details. Motivated by the understanding that global MDPs based regularisation approaches do not respect local variances of OF statistics, a novel spatially adaptive regularisation scheme is proposed. More specifically, the primary contribution of this chapter is a sparsity&non-sparsity constraints based prior-adaptive regularisation approach, that can be applied to general motion spaces, such as motion gradient or curvature spaces. The contribution of this chapter also includes the development of an Iteratively Reweighted Least Squares (IRLS) and Generalised Cross Validation (GCV) based strategy

that can simultaneously optimise the solutions for the flow field as well as the hyperparameter[1] fields.

8.1 Introduction

8.1.1 Background

Motion regularisation strategies are of crucial importance to OF estimation. Since OF estimation is an under-constrained problem, prior knowledge about the solution should be employed, so as to make the problem feasible. Recent advances in regularisation approaches have emphasised the three important aspects that have enabled significant improvements in OF estimation:

- Deciding motion integration regions, by taking advantage of the progress in modern low- or mid-level image segmentation approaches.

- Looking for improvements or replacements for existing motion spaces (ranging from original motion spaces to transformed spaces), or exploring more sophisticated spaces (e.g., by training).

- Choosing more suitable MDPs that can better fit/describe the statistics of a particular motion space.

In contrast to Chapter 4 that contributes a new non-local motion integration strategy, the current chapter concentrates on the MDP topic.

Existing OF regularisation schemes inherently assume that the OF signal, after projection into a considered motion space, has a particular structure, that complies with a pre-assumed MDP. By choosing a TV based smoothness regularity, the sparsity prior is to be enforced during the motion estimation process. This encourages piecewise constant motion flows, so, slowly fluctuating motion signals are flattened, while some high-frequency signals are preserved.

---

[1]In Bayesian statistics, a hyperparameter is a parameter of a prior distribution; the term is used to distinguish them from parameters of the model for the system under analysis.

In contrast, Tikhonov based regularities help to restore globally smooth motion fields. This is due to the fact that the associated MDPs (for Tikhonov based penalisers) are Gaussian distributions, so they are tolerant about low-frequency signals. However, by uniformly setting spatially invariant MDPs, it does not respect that, different local regions may have differing statistical distributions of motion flows, which may have large variances. Since imposing a global MDP in regularising OF estimation is in theory not appropriate, the main purpose of this chapter is to investigate the application of spatially adaptive MDPs in OF estimation.

8.1.2 Organisation

This chapter is organised as follows: in Section 2, most related works are reviewed and compared. Section 3 describes the mathematical formulation of the proposed new regularisation strategy, which is integrated into a unified OF framework. Due to the introduced hyperparameters that need to be simultaneously estimated with the OF vectors, the minimisation of the proposed objective functional is very challenging. So, implementation details are provided in Section 4. Specifically, a novel strategy is developed, based on the Mean-Field approximation theory, the IRLS method and the GCV technique, in order to solve the proposed, energy-minimisation functional with non only the OF field but also the hyperparameter fields as unknowns.

8.2 Related Works

The proposed approach is a TV&Tikhonov based spatially variant MDP that is to simultaneously model sparse and non-sparse motion statistical information in the OF field. This section aims to review the related regularisation approaches, and also clarify the research motivations.

8.2.1 MDPs for Gradient based Regularities

According to the survey in Chapter 3, the Tikhonov and the TV based penalisation functions are most common in OF regularisation schemes. When these two types of penalisation functions both collaborate with the traditional first-order derivative (f.o.d) based motion spaces (referred to as gradient spaces), the induced two types of regularities (briefly denoted as Tikhonov&f.o.d and TV&f.o.d regularities) are respectively good at recovering globally smooth flows and piecewise constant flows. This gives rise to a question: used in conjunction with the same motion space, why do TV and Tikhonov based regularities have different motion estimation capabilities?

This is due to the reason that: the TV penaliser encourages the motion gradient field to obey the Laplacian distribution. So, most of the values in the gradient space should be zero, and at the same time, only a small percentage of large values is allowed. Because of that, the TV&f.o.d type of regularities are effective in recovering piecewise constant flows, and simultaneously permitting motion discontinuities. However, TV&f.o.d regularities are overly-restricted for estimating motion flows that contain denser details of small or large gradients in the OF field (i.e. mid-frequency signals are penalised). In the case of applying the Tikhonov&f.o.d penaliser (also combined with the f.o.d based space): the projected motion-gradient signals are encouraged to obey the Gaussian distribution. So, non-sparse structures (such as small or large gradients) in the OF field can be kept. The main weakness of the Tikhonov&f.o.d regularity provides insufficient constraints for recovering sparse motion flows. Over-smoothed flow fields are usually resulted, where discontinuities or steep gradients are expected. Unfortunately, sparse and denser motion gradient signals usually coexist in the same OF field, it is thus necessary to design a regularisation scheme that supports spatially variant, statistical distributions of local regions in the gradient space.

8.2.2 MDPs for Curvature based Regularities

Recently, second-order derivatives (s.o.ds) based regularisation becomes popular in OF estimation as well as in the closely related stereo matching domain

(see the works (Ranftl et al., 2012, Yuan et al., 2009, Trobin et al., 2008), and (Woodford et al., 2009)). This is due to the fact that some types of s.o.ds based motion spaces provide efficient ways to describe piecewise affine transformations.

According to the literature survey in Chapter 3, s.o.ds based spaces are usually used in conjunction with TV penalisers in OF regularisation, such as in the works of Trobin et al. (2008) and Yuan et al. (2009). Because of the exploited TV function, the sparsity prior is enforced on the s.o.ds induced motion curvature spaces. The regularisation approaches of Trobin et al. (2008) and Yuan et al. (2009) (briefly denoted as TV&s.o.d or referred to as sparse curvature regularities) can achieve good performance in recovering piecewise affine flows. The assumption that sparse structures exist in motion curvature spaces complies well with the high-level understanding that the curvature space of a piecewise affine motion field has only a small percentage of high curvature signals, and most of the curvature magnitudes are zero. However, for describing motion fields with non-rigid deformation[2], sparse curvature regularities might be overly-restricted.

For dealing with non-rigid motion estimation: Li et al. (2013) recently proposed a Laplacian Mesh Deformation (LMD) based regularity, where the mean curvature of the motion surface is enforced to be as small as possible. In the work of Kadri-Harouna et al. (2013), a s.o.d based formulation combined with the Tikhonov penaliser is designed to encode a divergence-free prior for estimating turbulent fluid motion. Apart from that, it is interesting to note that, in the image registration world, the Tikhonov&s.o.d type of regularisation (usually referred to as curvature based registration) has, for a long time, been applied to non-rigid flow estimation, and has generated satisfactory results (Fischer and Modersitzki, 2003, Henn, 2006, Köstler et al., 2008, Sotiras et al., 2013). The similarity of the regularisation approaches (Fischer and Modersitzki, 2003, Köstler et al., 2008, Li et al., 2013, Kadri-Harouna et al., 2013) is that they all combine a s.o.d based motion space and the Tikhonov penalisation, for directly modelling and regularising non-rigid flows. One problem of those regularisation approaches is that, by exploiting the Tikhonov penalisation, it actually assumes that the associated motion curvature space is a

---

[2]In this work, non-rigid motion refers to the non-affine type of non-rigid motion.

denser or non-sparse field. Consequently, regularisation approaches based on this assumption enforce a globally smooth field and penalise sparse structures and motion discontinuities. Therefore, the non-sparse curvature regularity is not suitable for the estimation of piecewise affine fields, or even piecewise non-rigid fields that have occasional motion discontinuities. However, in a general flow field of a real-world scene, different motion patterns usually coexist. The investigation of s.o.ds based regularisation approaches also leads to the understanding that a spatially variant MDP is required.

8.2.3 Automated Estimation of MDPs

According to the survey in Chapter 3 as well as the above discussion, although there are various motion spaces based regularities (e.g., derivatives based and WTs based) in the literature, it is surprising to find that the distribution prior is usually pre-set and globally fixed. This assumption is however not proper for OF fields of real-world datasets, since variant motion statistics might exist within different local or nonlocal regions of the same image, and also take place in different datasets.

To the best of this author's knowledge, the idea of automatically selecting or spatially adapting the MDPs has emerged very lately in the OF estimation domain. Héas et al. (2012) proposed a method that automatically selects distribution priors for different datasets. Simpson et al. (2013) very recently suggest to decompose a Gaussian kernel based prior into several components. By allowing the weights of the Gaussian components to vary spatially, they have enabled a spatially adaptive regularisation for image registration problems. Another related work is (Chantas et al., 2014) that enables spatially adaptive motion-regularisation models for OF estimation.

The main difference between (Héas et al., 2012) and the proposed approach is that, rather than fixing a global distribution prior, spatially varied distribution priors are considered in the proposed work. The adaptive regularisation approach of Simpson et al. (2013) relies on a Gaussian distribution based prior. By comparing the proposed approach with (Chantas et al., 2014), the following key differences are noted:

- The main purpose of Chantas et al. (2014) is to reform the traditional Horn-Schunck OF algorithm into a variational Bayesian framework. The proposed strategy in this section follows the existing regularisation approaches in deterministic OF frameworks. Specifically, the proposed approach is inspired by: 1) the work of Krähenbühl and Koltun (2012), who demonstrate that many existing, well-known smoothness penalisers can be uniformly approximated by mixtures of exponential distribution kernels; and 2) Cho et al. (2010), who exploit a generalised Gaussian distribution model to simultaneously model sparse and non-sparse gradient distributions in local image patches, for the task of image restoration. Specifically, the idea of Cho et al. (2010) particularly inspired this work in the exploration of sparse and non-sparse signal distributions.

- Another key difference is that: by employing the Student's t-distribution (Hill, 1970) to model the motion distribution, the approach of Chantas et al. (2014) actually enforces Gaussian distributions based priors for each local region, inside the considered motion-gradient space. Their mechanism of spatially adaptive regularisation is enabled by spatially varying the variance value for the associated Gaussian distribution. Since the estimated value for the *degrees of freedom parameter* (a global parameter, associated with the Student's t distribution) is always small, a considerable rate of Gaussian kernels with large variances are allowed. By this means, in some local regions, large motion discontinuities can be considered as "outliers". In its essence, the regularisation scheme of Chantas et al. (2014) enforces non-sparse MDPs everywhere. So their approach is not efficient where a sparsity prior is preferred. By contrast, the work in this chapter takes into account both Gaussian and Laplacian based MDPs. By doing so, spatially variant non-sparse or sparse statistics that might coexist in the motion gradient/curvature spaces can be more properly modelled.

- Compared with (Chantas et al., 2014) that imposes the regularisation constraint only in the motion gradient field, the current work is also inspired by the understanding that, in more complex motion spaces, such as higher-order motion spaces, spatially variant MDPs are also required and should be made possible.

In summary, is has been overlooked in the literature that spatially variant MDPs are very important for OF estimation. To this end, this section proposes a spatially adaptive MDP, in order to flexibly enable sparse and non-sparse signal structures in different local or non-local regions in the considered motion space, for the purpose of improving the accuracy of OF estimation.

8.3 A Sparsity&Non-Sparsity Constraints based Prior-Adaptive Regularity

Recall the generalised Gaussian model (GGM) based, generalised regularisation term in Section 3.6,

$$E_S^0(\mathbf{v}, L, \eta, \sigma) = \sum_{\mathbf{x}} |\frac{L\mathbf{v}}{\sigma}|^{\eta} \tag{8.1}$$

where L represents the operator that transforms the field $\mathbf{v}$ into a desired domain or space. The hyperparameters η and σ are the shape parameters of local latent distributions in the considered motion space. $|\cdot|$ denotes the total variation. Note that this formulation treats L, η and σ as unknowns.

Separately treating L or (η, σ) as unknowns corresponds to two different problems: the automated estimation of motion spaces, and the automated estimation of MDPs. It can be anticipated that, if both groups of parameters L and (η, σ) are to be sought together with the field $\mathbf{v}$, a very difficult, or technically untackable energy-minimisation issue will result. As far as it is known, the potential underlying connections between these two problems have not been investigated in the literature. Based on the knowledge from the compressive sensing field, for a global region or different local regions in a targeted OF field, there should be optimal motion-space candidates that are able to more compactly describe the projected signals. Since it is not clear yet how to find optimal motion-space candidates for global/local motion fields, in this work, commonly adopted motion spaces are considered for L, such as derivatives based spaces and wavelet transformations (WTs) induced spaces. This means that L in Equation (8.1) will be pre-selected. By doing so, the current research focus narrows down on the topic of automatically estimating spatially variant MDPs.

The proposed, spatially adaptive MDPs based regularisation term (with L as known) is thus formulated as,

$$E_S(\mathbf{v}, \eta, \sigma) = \sum_{\mathbf{x}} \frac{|L\mathbf{v}|^{\eta}}{\sigma} \tag{8.2}$$

where $\eta \in (0, 2]$ determines the peakiness of the associated motion-distribution pdf, and $\sigma > 0$ is the width of the pdf (see corresponding explanations in Section 3.6). The larger the value of η, the flatter the pdf shape for the MDP. Smaller η means a more peaked shape. Due to the design, the L1-norm and the Tikhonov based regularities are two particular cases when setting $\eta = 1$ and $\eta = 2$ in Equation (8.2). Note that the motion-space projector L is pre-chosen for this regularisation term. General examples for L can be gradient operator ∇, s.o.ds based or WTs based operators.

Further, by combining the proposed regularisation term with a standard data-conservation term, the following OF energy functional is formulated as,

$$E(\mathbf{v}, \eta, \sigma) = \gamma E_D(\mathbf{v}) + E_S(\mathbf{v}, \eta, \sigma) \tag{8.3}$$

with

$$E_D(\mathbf{v}) = \sum_{\mathbf{x}} (I_x(\mathbf{x})u(\mathbf{x}) + I_y(\mathbf{x})v(\mathbf{x}) + I_t(\mathbf{x}))^2 \tag{8.4}$$

where $\mathbf{v} = (u, v)^T$, $I_x = \frac{\partial I_2}{\partial x}, I_y = \frac{\partial I_2}{\partial y}$ and $I_t = I_2 - I_1$. So, there are four fields of unknowns in the objective functional, namely u, v, η, and σ. Note that the data term can use many other forms (see Section 2.1.1 for data term candidates).

Gaussian priors (corresponding to $\eta = 2$ in the GGM) are good for modelling mid-frequency signals (either small or relatively large signals), and sparse structures in the motion field can be modelled well by Laplacian distributions (corresponding to $\eta = 1$ in the GGM) as well as hyperLaplacian distributions ($0.5 \leq \eta \leq 0.8$) (Krishnan and Fergus, 2009). Since introducing hyperLaplacian based penalisers (into Equation (8.3)) leads to highly non-convex energy-minimisation problems, in practice, the Laplacian-distribution based penaliser is usually exploited as a replacement for hyperLaplacian-distribution based ones. So, a discrete value set $\{\eta = 1, 2\}$ is employed. This means

that only Laplacian and Gaussian distributions are considered in the proposed algorithm.

Additionally, in the case of $\eta(\mathbf{x}) = 2$, the enforced regularisation constraint for a local motion-region around pixel $\mathbf{x}$ is a Gaussian-distribution based prior. With the larger the value of $\tau(\mathbf{x})$, the pdf shape of the associated Gaussian distribution has a smaller variance. So, most of the gradient signals are encouraged to be small in the local region around the pixel $\mathbf{x}$. Motion regions with small τ values correspond to allowing a considerable rate of relatively large gradients. The Gaussian distribution based MDP with a relatively small variance is very common in the regularisation literature. This type of MDPs, due to the tolerance of small gradients, can efficiently recover piecewise smooth signals. A Gaussian distribution with a large variance (i.e. a small τ) leads to a MDP that permits non-sparse structures of relatively large gradients. As far as it is known, the latter type of MDPs has not been exploited in OF regularisation. It is only recently, in the field of image restoration, Cho et al. (2010) proposed a regularisation approach that involves MDPs with large variances of Gaussian distributions, for restoring fractal-like textures, such as tree leaves and grass.

The smoothness penaliser in Equation (8.2) takes into account both sparse and non-sparse MDPs during motion estimation. For each image pixel in the considered motion field, apart from calculating the associated OF vector, a theoretically optimal MDP is also estimated. Therefore, the proposed strategy is referred to as the sparsity&non-sparsity constraints based prior-adaptive regularisation approach.

8.4 An Approximation Solution for the Objective Functional

Minimising the proposed energy functional in Equation (8.3) involves simultaneously computing the field $\mathbf{v}$ and the hyperparameter maps of η and σ. So, the optimisation of the proposed functional is very challenging. Because of this, detailed implementation procedures are described in this section.

Different strategies have been proposed in the literature to estimate hyperparameters, such as (generalised) cross-validation (GCV) (Wahba, 1977, Foroosh, 2005, Favati et al., 2014), Mean-Field approximation (MFA) (Geiger and Girosi, 1991, Xu et al., 2012), maximum a posterior (MAP), etc. Inspired by the works of Geiger and Girosi (1991) and Xu et al. (2012), a MFA based approach is employed to marginalise out the hyperparameter η to simplify the objective energy functional, at the first stage. The hyperparamter σ is dealt with by a GCV based approach at another separate stage. The estimation of the $\mathbf{v}$ and σ fields are simultaneously performed according to the Iteratively Reweighted Least Squares method.

Before going to the next section, since η takes a discrete value set $\{1, 2\}$, a binary field α is introduced into the regularisation term (Equation (8.2)). This will transform the regulariser into a more comprehensive form,

$$\tilde{E}_S(\mathbf{v}, \alpha, \tau) = \sum_{\mathbf{x}} \{\alpha(\mathbf{x})\tau(\mathbf{x})|\nabla\mathbf{v}(\mathbf{x})|^1 + (1 - \alpha(\mathbf{x}))\tau(\mathbf{x})|\nabla\mathbf{v}(\mathbf{x})|^2\} \quad (8.5)$$

with $\tau(\mathbf{x}) = \frac{1}{\sigma(\mathbf{x})}$. The binary variable $\alpha(\mathbf{x})$ works as a sparsity&non-sparsity switcher that guides whether the TV or the Tikhonov based penaliser will be chosen in the local/non-local region around $\mathbf{x}$. Correspondingly, the objective functional is changed to,

$$E(\mathbf{v}, \alpha, \tau) = \gamma E_D(\mathbf{v}) + \tilde{E}_S(\mathbf{v}, \alpha, \tau) \quad (8.6)$$

8.4.1 Averaging Out the Sparsity&Non-Sparsity Switching Process

According to the Clifford-Hammersley theorem, the probability of a particular state of the system is given by

$$P(\mathbf{v}, \alpha, \tau) = \frac{1}{Z}e^{-\beta E(\mathbf{v}, \alpha, \tau)} \quad (8.7)$$

where β is the inverse temperature, and Z is the normalising factor called the partition function, defined as

$$Z = \sum_{\{\mathbf{v}\}} \sum_{\{\alpha\}} \sum_{\{\tau\}} e^{-\beta E(\mathbf{v}, \alpha, \tau)} \quad (8.8)$$

where $\{\mathbf{v}\}$ denotes all the possible configurations of $\mathbf{v}$ of the system. $\{\alpha\}$ represents the set of values that the hyperparameter α can take, i.e., $\{\alpha\} = \{0, 1\}$, and $\{\tau\}$ is the set of candidate values for τ (the inverse of the variance parameter σ). The computation of the partition function requires the evaluation of a multi-dimensional integral that can not be explicitly solved, due to the variable interaction inside $E(\mathbf{v})$. So, an approximate solution is sought by making use of MFA. Note that MFA is a general tool used in statistical mechanics. It consists in substituting the interaction among the fields at different locations by the interaction of the field at each site with the mean field value at different locations (Geiger and Girosi, 1991).

The first task is to average out the sparsity&non-sparsity switching process through MFA, i.e. marginalising out the $\alpha(\mathbf{x})$ parameter from the system (Equation (8.6)). For the moment, τ is assumed to be known in advance. Z can then be re-written as

$$Z = \sum_{\{\mathbf{v}\}} \sum_{\alpha = \{0,1\}} e^{-\beta E(\mathbf{v}, \eta)} \tag{8.9}$$

The sum is first computed over all possible α values, which yields,

$$
\begin{aligned}
Z &= \sum_{\{\mathbf{v}\}} e^{-\beta \gamma E_D(\mathbf{v})} \sum_{\alpha\{=0,1\}} e^{-\beta \sum_{\mathbf{x}} \{\alpha(\mathbf{x})\tau(\mathbf{x})|\nabla\mathbf{v}(\mathbf{x})|^1 + (1-\alpha(\mathbf{x}))\tau(\mathbf{x})|\nabla\mathbf{v}(\mathbf{x})|^2\}} \\
&= \sum_{\{\mathbf{v}\}} e^{-\beta \gamma E_D(\mathbf{v})} \prod_{\mathbf{x}} (e^{-\beta R_1(\mathbf{v}(\mathbf{x}))} + e^{-\beta R_2(\mathbf{v}(\mathbf{x}))}) \\
&= \sum_{\{\mathbf{v}\}} e^{-\beta \{\gamma E_D(\mathbf{v}) - \sum_{\mathbf{x}} \frac{1}{\beta} \ln(e^{-\beta R_1(\mathbf{v}(\mathbf{x}))} + e^{-\beta R_2(\mathbf{v}(\mathbf{x}))})\}}
\end{aligned}
\tag{8.10}
$$

where $R_1(\mathbf{v}(\mathbf{x})) = \tau(\mathbf{x})|\nabla\mathbf{v}(\mathbf{x})|$ and $R_2(\mathbf{v}(\mathbf{x})) = \tau(\mathbf{x})|\nabla\mathbf{v}(\mathbf{x})|^2$. Then, either using the strategy of Geiger and Girosi (1991), or simply referring to the saddle point approximation approach, a same approximated result for Z is resulted,

$$Z \approx e^{-\beta\{\gamma E_D(\mathbf{v}) - \sum_{\mathbf{x}} \frac{1}{\beta} \ln(e^{-\beta R_1(\mathbf{v}(\mathbf{x}))} + e^{-\beta R_2(\mathbf{v}(\mathbf{x}))})\}} \tag{8.11}$$

Because of the above adaptation, the sparsity&non-sparsity switching process can thus be averaged out in the partition function.

According to the transformed simpler form of Z, the effective energy functional is yielded,

$$E^{eff}(\mathbf{v}) = \gamma E_D(\mathbf{v}) - \sum_{\mathbf{x}} \frac{1}{\beta} \ln(e^{-\beta R_1(\mathbf{v}(\mathbf{x}))} + e^{-\beta R_2(\mathbf{v}(\mathbf{x}))}) \tag{8.12}$$

$$E_{Reg}^{eff}(\mathbf{v}) = \sum_{\mathbf{x}} \frac{1}{\beta} \ln(e^{-\beta R_1(\mathbf{v}(\mathbf{x}))} + e^{-\beta R_2(\mathbf{v}(\mathbf{x}))}) \tag{8.13}$$

Once the Mean-Field (MF) approximation of the flow $\mathbf{v}$ is obtained, the approximation of Z can be achievable. As shown by Geiger and Girosi (1991), the MF approximation of the switching process α can be computed using

$$\bar{\alpha}(\mathbf{x}) = -\frac{1}{\beta} \frac{\partial \ln Z}{\partial G(\mathbf{x})} = \frac{1}{1 + e^{\beta(R_1(\mathbf{v}(\mathbf{x})) - R_2(\mathbf{v}(\mathbf{x})))}} \tag{8.14}$$

where $G(\mathbf{x})$ is a coefficient, set to $R_1(\mathbf{v}(\mathbf{x})) - R_2(\mathbf{v}(\mathbf{x}))$. So given β and the approximated flow $\mathbf{v}$, the MF approximation of $\alpha(\mathbf{x})$, i.e., $\bar{\alpha}(\mathbf{x})$, can be obtained. The effect of $\bar{\alpha}(\mathbf{x})$ equates that of $\alpha(\mathbf{x})$.

To summarise, the underlying idea of MFA is to first eliminate the MDP's degree of freedom from Z. By doing so, the effect of the interaction between the hyperparameter α (or η) with the field $\mathbf{v}$ can be simulated by an "effective potential", that depends only upon $\mathbf{v}$ and τ.

8.4.2 IRLS based Flow-Field Estimation

Based on the MFA scheme, the field of $\bar{\alpha}$ is obtained, i.e. η can be determined. Therefore, the objective functional with the already-known hyperparameter η can be written as

$$E(\mathbf{v}, \tau) = \gamma \sum_{\mathbf{x}} (I_x(\mathbf{x}) u(\mathbf{x}) + I_y(\mathbf{x}) v(\mathbf{x}) + I_t(\mathbf{x}))^2 + \sum_{\mathbf{x}} \tau(\mathbf{x}) |(L\mathbf{v})(\mathbf{x})|^{\eta(\mathbf{x})} \tag{8.15}$$

At this stage, still assuming that the field of τ is known, a so-called Iteratively Reweighted Least Squares (IRLS) technique will be employed to find the solution of $\mathbf{v}$. This first requires the representation of the OF objective functional (Equation (8.15)) with a more compact, matrix based formulation,

$$E(U, \tau) = \gamma |HU - b|^2 + \tau |LU|^{\eta} \tag{8.16}$$

with U, H, L defined as,

$$H = [\mathbf{I}_x \ \mathbf{I}_y] \in \mathbf{R}^{N \times 2N},$$

$$U = \begin{pmatrix} u_1 & u_2 & \cdots & u_N & v_1 & v_2 & \cdots & v_N \end{pmatrix}^T \in \mathbf{R}^{2N \times 1} \tag{8.17}$$

$$b = \begin{pmatrix} [I_t]_1 & [I_t]_2 & \cdots & [I_t]_N \end{pmatrix}^T \in \mathbf{R}^{N \times 1}$$

where N is the total number of pixels in the image, $\mathbf{I}_x = \mathrm{diag}(I_x), \mathbf{I}_y = \mathrm{diag}(I_y)$, and $[I_t]_i$ stands for the value in the linear index location i in I_t. L represents the matrix that could correspond to a f.o.d or s.o.d based space transformator. For example, taking $x-$ or $y-$ directional derivative filters as examples, LU denotes the convolution of U with the mask $[0 - 1\ 1]$.

Inspired by Wohlberg and Rodriguez (2007) and Daubechies et al. (2010), the optimisation processes of two special cases (related to Equation (8.16)) are examined in detail:

- When all values in the η field are equal to 2 (corresponding to using the Tikhonov based regularity everywhere in the motion field), the solution of $\mathbf{v}$ in Equation (8.16) is given by,

$$U_1 = (H^T H + \tau L^T Q_1 L)^{-1} H^T b \qquad (8.18)$$

  where Q_1 is an identity matrix $\mathbb{I}$. This equation provides an analytic solution of U_1, for OF objective functionals[3] that use the spatially invariant Tikhonov penaliser in the smoothness term. The value of U_1 can be determined using direct solvers for small-scale problems and iterative solvers, for example a conjugate gradient (CG) method, for large-scale problems.

- When all values in the η field are equal to 1, a similar equation to (8.18) can be obtained,

$$U_2 = (H^T H + \tau L^T Q_2 L)^{-1} H^T b \qquad (8.19)$$

  with Q_2,

$$Q_2 = \mathrm{diag}_{i=1,2,...,N}\Big(\frac{1}{\sqrt{[LU_2]_i^2 + \epsilon}}\Big) \qquad (8.20)$$

where $i \in \{1, 2, ..., N\}$ represents the linear index for the OF vectors involved. $[LU_2]_i$ stands for the value in the linear index location i in the matrix LU_2. ϵ is a small positive value. This form of U_2 solution corresponds to where a TV or L1-norm based spatially invariant MDP is employed. Since Q_2 depends

---

[3]All the considered OF objective functionals use the same data term (as in Equation (8.16), except where noted.

on U_2, U_2 elements exist on both sides of Equation (8.19). This is different from that, in Equation (8.18), there is no U_1 element on the right hand side of the equation. Therefore, Equation (8.19) can not be directly solved. However, thanks to the works of Wohlberg and Rodriguez (2007), Daubechies et al. (2010) and Liu (2009), IRLS can be used to estimate U_2 for Equation (8.19). Table 8.1 summarises the processes of IRLS.

Table 8.1: The IRLS method to find the solution for Equation (8.19)

Step 1:	set $k = 0$, $Q^0 = \mathbb{I}$, $U_2^0 = 0$
Step 2:	solve $U_2^{k+1} = (H^T H + \tau L^T Q_2^k L)^{-1} H^T b$
Step 3:	define $Q_2^{k+1} = \mathrm{diag}_{i=1,2,\dots,N} \left(\frac{1}{\sqrt{[LU_2^{k+1}]_i^2 + \epsilon}} \right)$
Step 4:	if U_2 converges, stop; else set $k = k + 1$ and go to Step 2

Returning to the original Equation (8.16) with spatially variant η values, an equation similar to (8.18) and (8.19) can be written as,

$$U = (H^T H + \tau L^T Q L)^{-1} H^T b \tag{8.21}$$

where Q, a diagonal matrix, is defined as,

$$Q_{ii} = \begin{cases} 1 & \text{if } \eta_i = 2, \\ \frac{1}{\sqrt{[LU]_i^2 + \epsilon}} & \text{if } \eta_i = 1. \end{cases} \tag{8.22}$$

where η_i represents the η value in the linear index location i. By analogy with how U_2 is optimised (according to IRLS), U in Equation (8.21), can be solved in an iterative matter, under the condition of pre-estimated τ and η fields.

With the assumption that τ values are known, Equation (8.15) can be also solved by the commonly used Euler-Lagrange method. The reason for choosing the IRLS based approach is three-fold:

- Different MDPs based penalisers, apart from the TV and the Tikhonov based, can be easily incorporated into the IRLS framework by changing the matrix Q, i.e. the weighting parameter matrix. Derivatives based motion spaces, or WTs induced spaces can be also easily exploited, corresponding to choosing a desirable L operator in Equation (8.21).

143

- An IRLS based approach provides a more straightforward way for estimating the τ field. Details will be present in the next section.

- As proved by Liu (2009), the IRLS based approach is identical to the variational method in minimising this type of energy functionals as proposed.

8.4.3 GCV and IRLS based Local Variance Estimation

The remaining problem is how to estimate τ. Regarding this issue, a strategy that combines GCV and IRLS is developed.

First, considering the problem of estimating a globally invariant τ value, using the GCV based approach amounts to minimising (Galatsanos and Katsaggelos, 1992)

$$A(\tau) = \frac{|b - HU|^2}{tr(\mathbb{I} - HUb^{-1})^2} \quad = \frac{|(\mathbb{I} - H(H^T H + \tau L^T Q L)^{-1} H^T)b|^2}{(tr(\mathbb{I} - H(H^T H + \tau L^T Q L)^{-1} H^T)^2} \quad (8.23)$$

with respect to τ, where $tr(\cdot)$ denotes the trace of the matrix. U is defined in Equation (8.21). Inspired by Galatsanos and Katsaggelos (1992), and Foroosh (2005), a simplified expression is to be sought for the minimiser of this GCV criterion. By letting $P = \mathbb{I} - H(H^T H + \tau L^T Q L)^{-1} H^T$, the GCV function can then be written as

$$A(\tau) = \frac{|Pb|^2}{(tr(P))^2} \quad (8.24)$$

Rearranging P leads to,

$$\begin{aligned} P &= \mathbb{I} - HL^{-1}L(H^T H + \tau L^T Q L)^{-1} L^T L^{-T} H^T \\ &= \mathbb{I} - HL^{-1}(L^{-T} H^T H L^{-1} + \tau L^{-T} L^T Q L L^{-1})^{-1} L^{-T} H^T \\ &= \mathbb{I} - K(K^T K + \tau Q)^{-1} K^T \\ &= \mathbb{I} - (\mathbb{I} + \tau K^{-T} Q K^{-1})^{-1} \end{aligned} \quad (8.25)$$

According to the matrix inversion lemma[4] (Tylavsky and Sohie, 1986), Equation (8.25) can be further transformed into,

$$P = \left(\mathbb{I} + \frac{1}{\tau} K Q^{-1} K^T\right)^{-1} \quad (8.26)$$

---

[4] $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$, for invertible A and C.

with $K = HL^{-1}$. Since $KQ^{-1}K^T$ is a symmetric matrix, the spectral decomposition of $KQ^{-1}K^T$ can be written as,

$$KQ^{-1}K^T = \Xi\Lambda\Xi^T = \sum_{j=1}^{n} \lambda_j \xi_j \xi_j^T \tag{8.27}$$

where ξ_js are the columns of Ξ that are a set of vectors orthogonal to each other, i.e. eigenvectors, and λ_js are the corresponding eigenvalues. n denotes the number of eigenvectors of Ξ, and it is associated with the number of pixels involved. For the estimation of a globally invariant τ value, information of all pixels inside the image will be used. Putting $KQ^{-1}K^T$ into Equation (8.26) gives:

$$P = \sum_{j=1}^{n} \frac{\tau}{\tau + \lambda_j} \xi_j \xi_j^T \tag{8.28}$$

With $\Xi^T b = \mathbf{z} = [z_1, z_2, ..., z_n]^T$, the GCV function can be rewritten as (see (Galatsanos and Katsaggelos, 1992)),

$$A(\tau) = \frac{\sum_{j=1}^{n} \left(\frac{\tau}{\tau + \lambda_j}\right)^2 z_j^2}{\left(\sum_{j=1}^{2} \frac{\tau}{\tau + \lambda_j}\right)^2} \tag{8.29}$$

In order to further simplify the GCV function of $A(\tau)$, an idea is borrowed from the approach of Foroosh (2005). Considering that most of the information in $KQ^{-1}K^T$ is carried by the first singular value. Therefore, $KQ^{-1}K^T$ is approximated by the value of $\lambda_1 \nu_1 \nu_1^T$. In practice, due to the particular construction of the matrix $KQ^{-1}K^T$, the largest singular value is, in most situations, orders of magnitude larger than the second singular value. Therefore,

$$A(\tau) \approx \frac{\left(\frac{\tau}{\tau + \lambda_1}\right)^2 z_1^2 + \sum_{j=2}^{n} z_j^2}{\left(n - 1 + \frac{\tau}{\tau + \lambda_1}\right)^2} \tag{8.30}$$

Based on the simplified expression of the GCV function, in order to minimise the value of $A(\tau)$, the equation is differentiated with respect to τ, and set to zero. By this it means, the optimal variance hyperparameter of τ is given by

$$\hat{\tau} = \frac{\lambda_1 \sum_{j=2}^{n} z_j^2}{(n-1)z_1^2 - \sum_{j=2}^{n} z_j^2} \tag{8.31}$$

Since the calculation for the above $A(\tau)$ involves the total number of pixels in the image, a globally constant τ value is obtained. While, it is straightforward

to consider the estimation of a spatially variant τ field. This can be accomplished by first segmenting the motion field into superpixel regions, and then estimating τ values locally.

Noting that, inside the work (Foroosh, 2005), the matrix Q is an identity matrix, since the Tikhonov penaliser is employed in the regularity. So, τ values in their case can be directly calculated without pre-estimating $\mathbf{v}$ or U. However, in the current work, Q (as defined in Equation (8.22)) contains the elements of $\mathbf{v}$. Direct estimation of the field $\hat{\tau}$ is thus infeasible. Recall that, due to the non-linearity of the original Equation (8.16), $\mathbf{v}$ can not be totally separated from other values/parameters in Equation (8.21). So, an iterative approach is used to optimise $\mathbf{v}$. By analogy, $\hat{\tau}$ can be dealt with by IRLS. $\mathbf{v}$ and $\hat{\tau}$ can then be estimated in an iterative manner. Implementation steps that simultaneously estimate $\mathbf{v}$ and the two hyperparameters are summarised in the next section.

8.4.4 Summary of Implementation Steps

The proposed energy-minimisation strategy is mainly based on IRLS, which iteratively estimates the field $\mathbf{v}$, and the two agent fields α and τ (respectively corresponding to the hyperparameters η and σ). Given temporarily estimated fields $\mathbf{v}^k$, α^k and τ^k at some mid-stage k in the whole energy-minimisation process, by fixing two of the three temporary fields, the third field at the $k+1$ stage can be estimated, according to the previously designed Equations (8.21), (8.14), or (8.31). For example, by fixing α^k and τ^k, $\mathbf{v}^{k+1}$ can be calculated, according to Equation (8.21).

Note that, to start the whole process, there is a necessity to get some initial estimates for $\mathbf{v}$, and τ and/or α. The initial field of $\mathbf{v}$, denoted as $\acute{\mathbf{v}}$, can be obtained by OF algorithms such as the Horn-Schrunk approach. The values of the initial τ and α fields, represented as $\acute{\tau}$ and $\acute{\alpha}$, can be set as the corresponding, spatially invariant values that are commonly used in the OF literature. See Table 8.2 that presents the proposed optimisation steps that simultaneously estimate the flow and the hyperparameter fields.

Table 8.2: The proposed iterative optimisation steps that simultaneously estimate the flow and the hyperparameter fields

Step 1:	Construct pyramids for both of the images, and set the initial pyramid level $l = 0$. Set $\mathbf{v}_0$ and τ_0 as down-sampled versions of the initially estimated fields $\acute{\mathbf{v}}$ and $\acute{\tau}$ at level 0.
Step 2:	Propagate flow field $\mathbf{v}_l$ and τ_l from the last pyramid layer to the current level $l + 1$, which yields two temporal fields $\breve{\mathbf{v}}_{l+1}$, and $\breve{\tau}_{l+1}$.
Step 3:	Take $\breve{\mathbf{v}}_{l+1}$ and $\breve{\tau}_{l+1}$ into Equation (8.14) to compute $\bar{\alpha}_{l+1}$.
Step 4:	Take $\bar{\alpha}_{l+1}$ and $\breve{\tau}_{l+1}$ into Equation (8.21) to compute $\mathbf{v}_{l+1}$.
Step 5:	Based on $\bar{\alpha}_{l+1}$ and $\mathbf{v}_{l+1}$, according to Equations (8.31) and (8.22), calculate τ_{l+1}.
Step 6:	If $l \neq n - 1$, where n is the pre-assumed total number of the pyramid layers, $l = l + 1$, and go to the second step.

To deal with the problem of large motion estimation, a pyramid framework of OF estimation is adopted. Based on the description above, the processing steps (as listed in Table 8.2) are iterated in each pyramid layer of the motion estimation process.

At this stage, the most common motion space, the f.o.d based space, is employed, since the current concentration is to investigate the importance of adaptive regularisation priors. Other motion spaces, such as curvature spaces, can also be considered by correspondingly changing the L operator in Equation (8.2).

Since the optimisation process needs to simultaneously estimate the OF field and the hyperparameter fields, this involves a highly non-convex energy minimisation problem. The algorithm thus has the problem of converging to local minima. In practice, graph-cut based discrete optimisation approaches can be considered to relieve the problem.

8.5   Conclusions

This chapter has proposed a sparsity&non-sparsity constraints based prior-adaptive regularisation approach. Since the objective OF energy functional, due to the involved hyperparameters, is particularly challenging to minimise, the chapter has also developed an IRLS&GCV based strategy that can simultaneously optimise the solutions for the flow field as well as the hyperparameter fields. The proposed regularisation algorithm and the new optimisation strategy are thus the two contributions in this chapter. Based on the well-grounded research motivation (as elaborated in both this chapter and Chapter 3), the algorithm design procedure, and the detailed introduction of the implementation approach, there are ample reasons to believe that the proposed prior-adaptive regularisation approaches has theoretical correctness and feasibility in practice.

Chapter 9

Conclusions and Future Work

The first aim of this chapter is to summarise the work and contributions of this thesis. Then, the publications that directly stem from this research work are presented. The final part of this chapter suggests potential directions that are worthy of future research.

9.1   Summary

This section aims to summarise the whole thesis. The main findings and contributions that are resulted from this research work are highlighted.

The first half part of this thesis concerns optic-flow (OF) based motion estimation. The ill-posed problem of OF estimation needs prior knowledge about the solution space, so as to make the problem feasible. Regularisation strategies are thus of crucial importance to the accuracy of OF estimation. Recent advances in regularisation approaches have emphasised three important aspects, namely exploring new motion integration strategies, looking for improvements or replacements of existing motion spaces, and investigating more suitable motion-distribution priors (MDPs) that can better fit or describe the statistics of a particular motion space. Motivated by this, by concentrating on the first and the third aspects, two motion regularisation approaches have been proposed. To be more concrete,

- An oriented geodesic distance (OGD) based non-local regularisation approach has been developed. At the heart of this approach is a novel pairwise-feature-affinity measurement. To the best of this author's knowledge, this is the first time that geodesic distances based non-local regularisation schemes have been investigated in the literature in the context of optic flow.

- A sparsity&non-sparsity constraints based prior-adaptive regularisation approach. This is mainly motivated by the fact that globally fixed MDPs based regularities do not respect local variances of OF statistics, and thus, a spatially variant MDP is necessary to be introduced to the OF regularisation field. In the meantime, considering that the involved OF energy functional is particularly challenging to minimise, a novel IRLS&GCV based strategy has been designed that is able to simultaneously optimise the solutions for the flow field as well as the hyperparameter fields.

The proposal of the two regularisation schemes, including the novel motioninte-gration strategy, and the spatially variant MDP based regularisation approach, has constituted the first core contribution of this thesis.

Moreover, a comprehensive literature survey on OF regularisation approaches has been provided. This has finally led to the design of a new generalised regularisation formulation. The remarkable advantage of the generalised regularisation scheme relies on the fact that it better respects the following facts:

- Local or non-local motion statistics may be different from the global distribution in the same motion field.

- Different datasets may need tailed choices of MDPs and/or motion spaces.

The detailed bibliography content and the revealed underlying connections, between different existing regularities, will potentially draw more researchers' attention to the exploration of new MDPs and motion spaces based regularisation approaches. This investigation is very promising to further advance the development of the OF field.

The second half part of this thesis focuses on the problem of cell tracking. Quantitative analysis on whole populations of cells, and identification of cell division events has great meaning for the study of normal/abnormal phenotypes and the effectiveness of drugs, and thus plays vital roles in the biomedical research domain. Driven by that, the computer vision domain has seen increasing attention in the design of automated tools for cellular biology researchers.

Compared with non-biomedical tracking tasks, inherent difficulties associated with cell tracking include dynamic backgrounds, cellular contrast changes, ambiguous boundaries, frequent intercellular interactions, the dramatic intensity change during cell division and death, etc. Regarding that, the second major contribution in this work is a model and motion based cell tracking framework, that is particularly designed for automatically tracking dense cell populations, over phase-contrast image sequences. The proposed system fully exploits the complementary properties from model and motion based cell tracking approaches. So, the shortcomings of the two worlds are both suppressed. The outstanding advantage of the proposed approach is reflected by the fact that, many challenging problems have been solved or relieved, including accurately segmenting and tracking partially overlapped cells, consistently tracking elongated cells, and the tracking of large displacement cells. It is worth noting that these problems are commonly considered to be difficult, and have not been well solved or are usually ignored in existing works. Within the new cell tracking system, there are two points worthy emphasising:

- The dynamic directional, gradient vector flow (DDGVF) technique is, for the first time, applied to segmenting and tracking dense cell populations. It has demonstrated superior performances in precisely outlining ambiguous cell boundaries, and correctly tracking cells in clustered environments.

- A novel strategy that seamlessly integrates the snake and the OF technique has been proposed, by incorporating a new soft external-energy term into the objective functional. The new energy term encourages model and motion based tracking results to be coherent to each other.

A concentrated literature survey on state-of-the-art cell tracking approaches

and systems has also been provided. Thanks to the indepth analysis of the survey, important new trends have been identified. This gives valuable insight into the topic of cell tracking, and will encourage future research.

In addition, another important contribution of this thesis is the development of a motion-occlusion analysis based automated cell-division detection approach. The proposed approach is quite different from the existing segmentation or training based methods. The major advantage of this approach is that dividing cells with a variety of division behaviour can be successfully detected, without the requirement of training assistances.

9.2   Publications Arising

The following peer-reviewed publications have arisen from this research.

1. Oriented Geodesic-Distance based Non-Local Optic Flow Approach, Sha Yu, Molloy D., IEEE Visual Communications and Image Processing (VCIP) 2013, 17- Nov-13 - 20-Nov-13, Kuching, Malaysia. **Awarded Student Travel Grant.**

2. Optic Flow based Occlusion Analysis for Cell Division Detection, Sha Yu, Molloy D., International Conference on Robotics, Vision, Signal Processing & Power Applications (ROVISP) 2013. 10-Nov-13 - 12-Nov-13, Penang, Malaysia.

3. Optic Flow Providing External Force for Active Contours in Visually Tracking Dense Cell Population, Sha Yu, Molloy D., 2011 Irish Machine Vision and Image Processing Conference (IMVIP2011), 07-SEP-11 - 09-SEP-11, Dublin, Ireland.

4. Automated Image-based Cell Tracking in Dense Cell Datasets, Sha Yu, Molloy D., 2010 Irish Machine Vision and Image Processing Conference (IMVIP2010), 08-SEP-10 - 10-SEP-10, University of Limerick. **Awarded Best Paper**.

5. Automated Cell Tracking using a Model-based Approach in Phase-Contrast

Dense Cell Datasets, Sha Yu, Molloy D., BioPhotonics and Imaging Conference (BioPIC 2010), Dublin, Ireland, 2010.

6. Automated Vision-based Tracking of Cell Movement and Deformation in Image Sequences, Sha Yu, Molloy D., 2010 China-Ireland International Conference on Information and Communications Technologies (CI-ICT2010), 10-OCT-10 - 11-OCT-10, Wuhan, China.

Apart from the aforementioned, there are three journal papers in preparation listed below,

1. A Sparsity&Non-Sparsity Constraints based Prior-Adaptive Regularisation Approach for Motion/Contour Estimation.

2. A Novel Model and Motion based Framework for Tracking Dense Cell Populations in Phase-Contrast Datasets with Automated Mitosis Detection.

3. Smoothness and Sparsity Priors based Optic-Flow Regularisation Approaches: Literature Survey and Generalisation.

9.3   Directions for Future Research

A first future plan is to fulfill the evaluation of the proposed prior-adaptive regularisation approach. Since motion fields of real-world images usually contain spatially varying statistics, the application of this approach is promising. Specifically, it is very interesting to improve the motion estimation accuracy for cellular images of dense populations, where spatially varying densities of cells usually lead to spatially variant motion patterns.

Secondly, while most of the concepts developed in this thesis have been experimentally evaluated, some future work, directly related to the proposed regularisation approaches, can be considered to further advance the ideas in this research domain. Specifically, the generalisation of the state-of-the-art OF regularisation approaches, and the proposed spatially adaptive regularisation

schemes open up a wide range of future directions for research. To name a few,

1. Considering that different MDPs based penalisation functions can be easily incorporated into the proposed prior-adaptive regularity, a natural suggestion is to introduce more types of MDPs, in addition to the Laplacian and Gaussian based priors, into the prior-adaptive regularisation approach. By this means, it is reasonable to expect further increased accuracies for OF estimation.

2. At the current stage, the most common motion space, i.e., the gradient based space, has been adopted in the proposed regularisation schemes. Other motion spaces, such as higher-order derivatives based, and WTs induced spaces, can also be straightforwardly employed. Specifically, the curvature space based regularities have been proven very useful in the estimation of either piecewise affine motion flows, or globally non-rigid motion fields. While the current literature has not seen a direct regularisation strategy that can effectively regularise rigid and non-rigid flows that simultaneously exist in the same motion field. Thanks to the findings in this thesis, a first possible way to achieve the purpose is to exploit the curvature space based prior-adaptive regularisation approach. The rational relies on the high-level understanding that: the curvature space, if projected by a piecewise affine flow field, has only a small percentage of high curvature signals, and most of the curvature magnitudes are zero. When transforming a non-rigid flow region into the curvature space, the resulting field will contain a denser or non-sparse structure of curvatures. A spatially adaptive, sparse&non-sparse curvature constraints based regularisation approach is thus a direction that is worthy of investigation.,

3. The proposed OGD based regularity is essentially a novel pairwise-feature-affinity measurement, that is based on an orientation-adaptive geodesic path strategy. So, it can be extended as a non-local image smoothing technique that has an advantage of edge preservation. In addition, since the accurate feature grouping capability of OGD has been demonstrated in boundary regions in motion signal fields, it is reasonable to expect that the non-local OGD strategy can be transplanted into the image

154

segmentation domain as well.

4. Since the regularisation approaches, that have been proposed or are planned as future works, involve the ideas of the non-local motion integration strategy, as well as higher-order derivatives based constraints, the computation requirement for the associated OF approaches will be extremely high. Apart from that, in current literature, the topic of the enhancement of computation efficiency has also attracted more attention. These factors make it very appealing to looking for efficient ways to optimise OF estimation processes.

Thirdly, there are other promising directions that have not been mentioned up to now. They are also conceived during the process of the review and generalisation of the state-of-the-art OF regularisation approaches.

1. Another possible way for regularising rigid and non-rigid motion flows that coexist in the same field. From the existing works that take into account the estimation of non-rigid motion, some carefully designed regularisation strategies have been proposed. For example, Xu et al. (2008) present a multi-stage strategy to first compute a piecewise affine motion field, and then re-calculate the flows in candidate non-rigid motion regions (that are selected by a pre-estimated, motion-estimation confidence map in the early stages). However, during the flow recalculation process, an approximated TV penaliser is utilised in the smoothness term. This still results in piecewise constant flows, and hence produces over-smoothed flows. Sun, Sudderth and Black (2010) and Sun et al. (2012) exploit a semi-parametric regularisation scheme, by encouraging the motion in image super-pixels or layers to be close to pre-estimated affine flows, whilst allowing deviations from the affine motion. The basic logic of this strategy considers that a non-rigid flow region consists of a piecewise affine layer and a "more textured" residual layer. Similar to (Xu et al., 2008), a TV-like penalisation function is employed. While, it can be understood that the residual layer (i.e. the residual between the non-rigid layer and the associated affine flow layer) cannot be properly modelled by sparsity-inducing regularities only. According to the

understanding that a non-rigid flow layer can be decomposed by a piece-wise affine flow layer and a residual layer, another potential approach to model a rigid and non-rigid, coexisting field is to combine a TV based curvature regularity and an adaptive sparsity&non-sparsity prior based gradient regularity. In theory, this also leads to a regularisation approach that can adaptively choose to recover piecewise affine or constant flows or non-rigid motion surfaces.

2. Integrating segmentation into sparsity priors based regularisation. Recently, sparsity-based motion priors have achieved encouraging results for motion estimation. These priors assume that the flow can be sparsely represented in certain domains. Compared with traditional smoothness constraints, one advocated advantage of sparsity-based constraints lies in the simultaneous handling of motion homogeneities and motion discontinuities (Shen and Wu, 2010, Jia et al., 2011). By doing so, the issue that is associated with selecting motion integration regions might be avoided. However, allowing crossing-boundary regularisation is actually a two-side sword. Due to occlusions that exist near motion boundaries, flow vectors that are constrained by the OF data-conservation term are usually erroneous. Simultaneous modelling of smooth and discontinuous motion, however, will allow incorrect estimation that is caused by occlusions on motion boundaries. Therefore, purely relying on the sparsity based regularisation may not lead to well-localised flow edges in motion boundaries. So, there arises a natural suggestion to combine segmentation based motion estimation and sparse priors based regularisation.

3. Investigating the pairwise asymmetric motion-coherence constraint. In existing local or non-local regularisation schemes, the motion-coherence constraint is usually symmetrically enforced between the two pixels involved. It is well accepted that motion estimation for textureless features needs to borrow information from more distinguishable features. However, as long as the pairwise motion-supporter mode is symmetrical between the two pixels, unreliable motion information (such as from outliers, textureless features, and even incomplete motion derived from edge features) can potentially blur or contaminate the should-be reliable motion vectors (i.e. texture features, and corner features).

Bibliography

A.J. Hand, T. Sun, D. B. D. H. and MacNeil, S. (2009), 'Automated tracking of migrating cells in phase-contrast video microscopy sequences using image registration', *Journal of Microscopy* **234**(November 2008), 62–79.

Alvarez, L., Deriche, R., Papadopoulo, T. and Sánchez, J. (2007), 'Symmetrical dense optical flow estimation with occlusions detection', *Int. J. Comput. Vision* **75**(3), 371–385.

Alvarez, L., E. J. L. M. . S. J. (1999), A pde model for computing the optical flow, *in* 'Proc. XVI congreso de ecuaciones diferenciales y aplicaciones', Las Palmas de Gran Canaria, Spain, pp. 1349–1356.

Ayvaci, A., Raptis, M. and Soatto, S. (2010), Occlusion detection and motion estimation with convex optimization, *in* 'Advances in Neural Information Processing Systems 23', NIPS '10, pp. 100–108.

Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M. J. and Szeliski, R. (2011), 'A database and evaluation methodology for optical flow', *Int. J. Comput. Vision* **92**(1), 1–31.

Barnes, C., Shechtman, E., Finkelstein, A. and Goldman, D. B. (2009), 'Patchmatch: A randomized correspondence algorithm for structural image editing', *ACM Trans. Graph.* **28**(3), 24:1–24:11.

Barron, J. L., Fleet, D. J. and Beauchemin, S. S. (1994), 'Performance of optical flow techniques', *Int. J. Comput. Vision* **12**(1), 43–77.

Benmansour, F. and Cohen, L. (2011), 'Tubular structure segmentation based on minimal path method and anisotropic enhancement', *Int. J. Comput. Vision* **92**(2), 192–210.

Berridge, M. J. (2012), 'Cell signalling biology'.
   URL: *http://www.biochemj.org/csb/*

Birchfield, S. T. and Pundlik, S. J. (2008), Joint tracking of features and edges, *in* 'Proceedings of the Conference on Computer Vision and Pattern Recognition', CVPR '08.

Black, M. J. and Anandan, P. (1996), 'The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields', *Computer Vision and Image Understanding* **63**, 75–104.

Black, M. J. and Jepson, A. D. (1996), 'Estimating optical flow in segmented images using variable-order parametric models with local deformations', *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(10), 972–986.

Blake, A. and Zisserman, A. (1987), *Visual reconstruction*, Vol. 2, MIT press Cambridge.

Bredies, K., Kunisch, K. and Pock, T. (2010), 'Total generalized variation', *SIAM J. Img. Sci.* **3**(3), 492–526.

Brox, T., Bruhn, A., Papenberg, N. and Weickert, J. (2004), High accuracy optical flow estimation based on a theory for warping, *in* 'Proceedings of the 8th European Conference on Computer Vision - Volume Part I', ECCV '04, pp. 25–36.

Brox, T., Bruhn, A. and Weickert, J. (2006), Variational motion segmentation with level sets, *in* 'Proceedings of the 9th European Conference on Computer Vision - Volume Part I', ECCV '06, pp. 471–483.

Brox, T. and Malik, J. (2011), 'Large displacement optical flow: Descriptor matching in variational motion estimation', *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(3), 500–513.

Bruhn, A., Weickert, J. and Schnörr, C. (2005), 'Lucas/kanade meets horn/schunck: Combining local and global optic flow methods', *Int. J. Comput. Vision* **61**(3), 211–231.

Cai, J.-F., Dong, B., Osher, S. and Shen, Z. (2012), 'Image restoration: Total variation, wavelet frames, and beyond', *Journal of the American Mathematical Society* **25**(4), 1033–1089.

Candès, E., Wakin, M. and Boyd, S. (2008), 'Enhancing sparsity by reweighted l1 minimization', *Journal of Fourier Analysis and Applications* **14**(5-6), 877–905.

Caselles, V., Kimmel, R. and Sapiro, G. (1997), 'Geodesic active contours', *Int. J. Comput. Vision* **22**(1), 61–79.

Chambolle, A. (2004), 'An algorithm for total variation minimization and applications', *J. Math. Imaging Vis.* **20**(1-2), 89–97.

Chan, T. F., Golub, G. H. and Mulet, P. (1999), 'A nonlinear primal-dual method for total variation-based image restoration', *SIAM J. Sci. Comput.* **20**(6), 1964–1977.

Chantas, G., Gkamas, T. and Nikou, C. (2014), 'Variational-bayes optical flow', *J. Math. Imaging Vis.* pp. 1–15.

Charmi, M. A., Derrode, S. and Ghorbel, F. (2008), 'Fourier-based geometric shape prior for snakes', *Pattern Recogn. Lett.* **29**, 897–904.

Chatterjee, R., Ghosh, M., Chowdhury, A. S. and Ray, N. (2013), 'Cell tracking in microscopic video using matching and linking of bipartite graphs', *Computer Methods and Programs in Biomedicine* **112**(3), 422–431.

Chen, Z., Jin, H., Lin, Z., Cohen, S. and Wu, Y. (2013), Large displacement optical flow from nearest neighbor fields, *in* 'Proceedings of the 2013 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '13, pp. 2443–2450.

Chen, Z., Wang, J. and Wu, Y. (2012), Decomposing and regularizing sparse/non-sparse components for motion field estimation, *in* 'Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '12, pp. 1776–1783.

Cheng, J. and Foo, S. W. (2006), 'Dynamic directional gradient vector flow for snakes', *IEEE Trans. Img. Proc.* **15**(6), 1563–1571.

Cheng, Y. (1995), 'Mean shift, mode seeking, and clustering', *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 790–799.

Chesnaud, C., Refregier, P. and Boulet, V. (1999), 'Statistical region snake-based segmentation adapted to different physical noise models', *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(11), 1145–1157.

Cho, T. S., Joshi, N., Zitnick, C. L., Kang, S. B., Szeliski, R. and Freeman, W. T. (2010), A content-aware image prior, *in* 'Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '10, pp. 169–176.

Cohen, L. D. (1991), 'On active contour models and balloons', *CVGIP: Image Underst.* **53**(2), 211–218.

Cohen, L. D. and Kimmel, R. (1997), 'Global minimum for active contour models: A minimal path approach', *Int. J. Comput. Vision* **24**(1), 57–78.

Cosgriff, R. (1960), 'Identification of shape', *Res. Foundation, Ohio State Univ., Columbus* **23**(3), 54–62.

Dalal, N. and Triggs, B. (2005), Histograms of oriented gradients for human detection, *in* 'Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 - Volume 01', CVPR '05, pp. 886–893.

Daubechies, I., Devore, R., Fornasier, M. and Güntürk, C. S. (2010), 'Iteratively reweighted least squares minimization for sparse recovery', *Comm. Pure Appl. Math* **63**, 1–38.

Debeir, O., Van Ham, P., Kiss, R. and Decaestecker, C. (2005), 'Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes.', *IEEE Trans. Med. Imaging* **24**(6), 697–711.

Dewan, M. A. A., Ahmad, M. O. and Swamy, M. N. S. (2011), 'Tracking biological cells in time-lapse microscopy: An adaptive technique combining motion and topological features', *IEEE Trans. Biomed. Engineering* **58**(6), 1637–1647.

Dong, W., Zhang, L., Shi, G. and Wu, X. (2011), 'Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization', *IEEE Trans. Img. Proc.* **20**(7), 1838–1857.

Donoho, D. (2006), 'Compressed sensing', *Information Theory, IEEE Transactions on* **52**(4), 1289–1306.

Farzinfar, M., Xue, Z. and Teoh, E. K. (2008), Joint parametric and non-parametric curve evolution for medical image segmentation, *in* 'Proceedings of the 10th European Conference on Computer Vision: Part I', ECCV '08, pp. 167–178.

Favati, P., Lotti, G., Menchi, O. and Romani, F. (2014), 'Generalized cross-validation applied to conjugate gradient for discrete ill-posed problems', *Applied Mathematics and Computation* **243**(0), 258 – 268.

Fischer, B. and Modersitzki, J. (2003), 'Curvature based image registration', *J. Math. Imaging Vis.* **18**(1), 81–85.

Fleet, D. J. and Weiss, Y. (2005), *Mathematical Models in Computer Vision: The Handbook (Optical Flow Estimation)*.

Foroosh, H. (2005), 'Pixelwise-adaptive blind optical flow assuming nonstationary statistics', *IEEE Trans. Img. Proc.* **14**(2), 222–230.

Forstner, W. (1986), 'A feature based correspondence algorithm for image matching', *Intl. Arch. of Photogrammetry and Remote Sensing* pp. 150–166.

Freedman, D. and Turek, M. W. (2005), Illumination-Invariant Tracking via Graph Cuts, *in* 'Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', Vol. 2 of *CVPR '05*, pp. 10–17.

Galatsanos, N. and Katsaggelos, A. (1992), 'Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation', *IEEE Trans. Img. Proc.* **1**(3), 322–336.

Geiger, D. and Girosi, F. (1991), 'Parallel and deterministic algorithms from mrfs: Surface reconstruction', *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(5), 401–412.

Gil, D. and Radeva, P. (2003), Curvature vector flow to assure convergent deformable models for shape modelling, *in* 'Energy Minimization Methods in Computer Vision and Pattern Recognition', Vol. 2683 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 357–372.

Goldstein, T. and Osher, S. (2009), 'The split bregman method for l1-regularized problems', *SIAM J. Img. Sci.* **2**(2), 323–343.

Gulshan, V., Rother, C., Criminisi, A., Blake, A. and Zisserman, A. (2010), Geodesic star convexity for interactive image segmentation, *in* 'Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '10, pp. 3129–3136.

Hadamard, J. (1923), *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Yale University Press.

Hamou, A. K. and El-Sakka, M. R. (2010), 'Optical flow active contours with primitive shape priors for echocardiography', *EURASIP J. Adv. Signal Process* **2010**, 9:2–9:2.

Han, J., Qi, F. and Shi, G. (2011), Gradient sparsity for piecewise continuous optical flow estimation, *in* 'in Proceedings of the IEEE International Conference on Image Processing', ICIP '11, pp. 2341–2344.

He, L. and Schaefer, S. (2013), 'Mesh denoising via l0 minimization', *ACM Trans. Graph.* **32**(4), 64:1–64:8.

Héas, P., Herzet, C. and Mémin, É. (2012), 'Bayesian inference of models and hyperparameters for robust optical-flow estimation', *IEEE Trans. Img. Proc.* **21**(4), 1437–1451.

Henn, S. (2006), 'A full curvature based algorithm for image registration', *J. Math. Imaging Vis.* **24**(2), 195–208.

Hill, G. W. (1970), 'Acm algorithm 395: Student's t-distribution', *Commun. ACM* **13**(10), 617–619.

Horn, B. K. P. and Schunck, B. G. (1981), 'Determining optical flow', *Artificial Intelligence* **17**, 185–203.

Huh, S. and Chen, M. (2011), Detection of mitosis within a stem cell population of high cell confluence in phase-contrast microscopy images, *in* 'Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition', CVPR '11, pp. 1033–1040.

Ince, S. and Konrad, J. (2008), 'Occlusion-aware optical flow estimation', *IEEE Trans. Img. Proc.* **17**(8), 1443–1451.

Jia, K., Wang, X. and Tang, X. (2011), Optical flow estimation using learned sparse model, *in* 'Proceedings of the International Conference on Computer Vision', ICCV '11, pp. 2391–2398.

Jiang, R. M., Crookes, D., Luo, N. and Davidson, M. W. (2010), 'Live-cell tracking using sift features in dic microscopic videos', *IEEE Trans. Biomed. Engineering* **57**(9), 2219–2228.

Jurrus, E., Tasdizen, T., Koshevoy, P., Fletcher, P. T., Hardy, M., bin Chien, C., Denk, W. and Whitaker, R. (2009), 'Axon tracking in serial block-face scanning electron microscopy', *Medical Image Analysis* **13**(1), 180–188.

Kaakinen, M., Huttunen, S., Paavolainen, L., Marjomaki, V., Heikkila, J. and Eklund, L. (2014), 'Automatic detection and analysis of cell motility in phase-contrast time-lapse images using a combination of maximally stable extremal regions and kalman filter approaches', *Journal of Microscopy* **253**(1), 65–78.

Kadri-Harouna, S., Dérian, P., Héas, P. and Mémin, E. (2013), 'Divergence-free wavelets and high order regularization.', *Int. J. Comput. Vis.* **103**(1), 80–99.

Kanade, T., Yin, Z., Bise, R., Huh, S., Eom, S. E., Sandbothe, M. and Chen, M. (2011), Cell image analysis: Algorithms, system and applications, *in* 'IEEE Workshop on Applications of Computer Vision', WACV '11.

Kass, M., Witkin, A. and Terzopoulos, D. (1988), 'Snakes: Active contour models', *Int. J. Comput. Vision* **1**(4), 321–331.

Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A. and Yezzi, A. (1995), Gradient flows and geometric active contour models, *in* 'Proceedings of the Fifth International Conference on Computer Vision', ICCV '95, IEEE Computer Society, Washington, DC, USA, pp. 810–815.

Kitagawa, G. (1987), 'Non-gaussian state-space modeling of nonstationary time series', *Journal of the American Statistical Association* **82**, 1032.

Krähenbühl, P. and Koltun, V. (2012), Efficient nonlocal regularization for optical flow, *in* 'Proceedings of the 12th European Conference on Computer Vision - Volume Part I', ECCV '12, pp. 356–369.

Krishnan, D. and Fergus, R. (2009), Fast image deconvolution using hyper-laplacian priors, *in* 'Adv. Neural Information Processing Systems', NIPS '09, Vancouver, British Columbia, Canada., pp. 1033–1041.

Köstler, H., Ruhnau, K. and Wienands, R. (2008), 'Multigrid solution of the optical flow system using a combined diffusion- and curvature-based regularizer', *Numerical Linear Algebra with Applications* **15**(2-3), 201–218.

Kwon, D., Lee, K. J., Yun, I. D. and Lee, S. U. (2010), Solving mrfs with higher-order smoothness priors using hierarchical gradient nodes, *in* 'in Proceedings of the Asian Conference on Computer Vision', ACCV '10, pp. 121–134.

Lee, K. J., Kwon, D., Yun, I. D. and Lee, S. U. (2010), Optical flow estimation with adaptive convolution kernel prior on discrete framework, *in* 'Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '10, pp. 2504–2511.

Lempitsky, V. S., Roth, S. and Rother, C. (2008), Fusionflow: Discrete-continuous optimization for optical flow estimation, *in* 'Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '08.

Li, B. and Acton, S. T. (2007), 'Active contour external force using vector field convolution for image segmentation', *IEEE Trans. Img. Proc.* **16**(8), 2096–2106.

Li, C., Liu, J. and Fox, M. D. (2005), Segmentation of edge preserving gradient vector flow: An approach toward automatically initializing and splitting of snakes, *in* 'Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 - Volume 01', CVPR '05, IEEE Computer Society, Washington, DC, USA, pp. 162–167.

Li, F., Zhou, X., Ma, J. and Wong, S. (2010), 'Multiple nuclei tracking using integer programming for quantitative cancer cell cycle analysis', *IEEE Trans. Med. Imaging* **29**(1), 96 –105.

Li, K., Miller, E. D., Chen, M., Kanade, T., Weiss, L. E. and Campbell, P. G. (2008), 'Cell population tracking and lineage construction with spatiotemporal context', *Medical Image Analysis* **12**(5), 546–566.

Li, W., Cosker, D., Brown, M. and Tang, R. (2013), Optical flow estimation using laplacian mesh energy, *in* 'Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition', CVPR '13, pp. 2435–2442.

Li, Y. and Osher, S. (2009), 'A new median formula with applications to pde based denoising', *Communications in Mathematical Sciences* **7**(3), 741–753.

Liu, C. (2009), *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*, Massachusetts Institute of Technology.
  URL: *http://people.csail.mit.edu/celiu/Thesis/CePhDThesis.pdf*

Liu, C., Freeman, W. T. and Adelson, E. H. (2006), Analysis of contour motions, *in* 'Advances in Neural Information Processing Systems', NIPS '06, pp. 913–920.

Lowe, D. G. (1999), Object recognition from local scale-invariant features, *in* 'Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2', ICCV '99, IEEE Computer Society, Washington, DC, USA, pp. 1150–1157.

Lucas, B. D. and Kanade, T. (1981), An iterative image registration technique with an application to stereo vision, *in* 'in Int. Joint Conference in Artificial Intelligence', pp. 674–679.

Mémin, E. and Pérez, P. (2002), 'Hierarchical estimation and segmentation of dense motion fields', *Int. J. Comput. Vision* **46**(2), 129–155.

Mohamed, S., Heller, K. A. and Ghahramani, Z. (2012), Evaluating bayesian and l1 approaches for sparse unsupervised learning, *in* 'Proceedings of the International Conference on Machine Learning', ICML '12.

Mohammad-Djafari, A. (2012), 'Bayesian approach with prior models which enforce sparsity in signal and image processing', *EURASIP J. Adv. Sig. Proc.* **2012**, 52.

Murray, D. and Buxton, B. F. (1987), 'Scene segmentation from visual motion using global optimization', *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-9**(2), 220–228.

Myronenko, A. (2010), *Non-rigid Image Registration: Regularization, Algorithms and Applications*, Oregon Health & Science University.
URL: *http://books.google.ie/books?id=uS3pZwEACAAJ*

Myronenko, A. and Song, X. B. (2009), 'Adaptive regularization of ill-posed problems: Application to non-rigid image registration', *CoRR* **abs/0906.3323**.

Nagel, H. H. (1990), Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow, *in* 'Proceedings of the First European Conference on Computer Vision', ECCV '90, pp. 139–148.

Nagel, H.-H. and Enkelmann, W. (1986), 'An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences', *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 565–593.

Nath, S. K., Palaniappan, K. and Bunyak, F. (2006), Cell segmentation using coupled level sets and graph-vertex coloring, *in* 'Proceedings of the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part I', MICCAI '06, pp. 101–108.

Nawaz, M. W., Bouzerdoum, A. and Phung, S. L. (2011), Optical flow estimation using sparse gradient representation, *in* 'in Proceedings of the IEEE International Conference on Image Processing', ICIP '11, pp. 2681–2684.

Nir, T., Bruckstein, A. M. and Kimmel, R. (2008), 'Over-parameterized variational optical flow', *Int. J. Comput. Vision* **76**(2), 205–216.

Onkarappa, N. and Sappa, A. D. (2013), Laplacian derivative based regularization for optical flow estimation in driving scenario, *in* 'Computer Analysis of Images and Patterns', CAIP '13, pp. 483–490.

Ottaviano, G. and Kohli, P. (2013), Compressible motion fields, *in* 'Proceedings of the 2013 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '13, pp. 2251–2258.

Padfield, D. R., Rittscher, J., Thomas, N. and Roysam, B. (2009), 'Spatio-temporal cell cycle phase analysis using level sets and fast marching methods', *Medical Image Analysis* **13**(1), 143–155.

Papenberg, N., Bruhn, A., Brox, T., Didas, S. and Weickert, J. (2006), 'Highly accurate optic flow computation with theoretically justified warping', *Int. J. Comput. Vision* **67**, 141–158.

Peyré, G., Péchaud, M., Keriven, R. and Cohen, L. D. (2010), 'Geodesic Methods in Computer Vision and Graphics', *Foundations and Trends in Computer Graphics and Vision* **5**, 197–397.

Pichon, E., Westin, C.-F. and Tannenbaum, A. R. (2005), A hamilton-jacobi-bellman approach to high angular resolution diffusion tractography, *in* 'Proceedings of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part I', MICCAI '05, pp. 180–187.

Quelhas, P., Mendonça, A. M. and Campilho, A. (2010), Optical flow based arabidopsis thaliana root meristem cell division detection, *in* 'Proceedings of the 7th international conference on Image Analysis and Recognition - Volume Part II', ICIAR '10, pp. 217–226.

Ranftl, R., Gehrig, S., Pock, T. and Bischof, H. (2012), Pushing the limits of stereo using variational stereo estimation, *in* 'Intelligent Vehicles Symposium', pp. 401–407.

Ray, N. and Acton, S. T. (2004), 'Motion gradient vector flow: an external force for tracking rolling leukocytes with shape and size constrained active contours', *IEEE Trans. Med. Imaging* **23**(12), 1466–1478.

Ray, N., Acton, S. T. and Ley, K. (2002), 'Tracking leukocytes in vivo with shape and size constrained active contours', *IEEE Trans. Med. Imaging* **21**(10), 1222–1235.

Ren, X. (2008), Local grouping for optical flow, *in* 'Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '08, pp. 1–8.

Rodriguez, M., Ali, S. and Kanade, T. (2009), Tracking in unstructured crowded scenes, *in* 'Proceedings of the International Conference on Computer Vision', ICCV '09, pp. 1389 –1396.

Roth, S. and Black, M. J. (2007), 'On the spatial statistics of optical flow', *Int. J. Comput. Vision* **74**(1), 33–50.

Rudin, L. I., Osher, S. and Fatemi, E. (1992*a*), 'Nonlinear total variation based noise removal algorithms', *Phys. D* **60**(1-4), 259–268.

Rudin, L. I., Osher, S. and Fatemi, E. (1992*b*), 'Nonlinear total variation based noise removal algorithms', *Phys. D* **60**(1-4), 259–268.

Seong, J.-K., Jeong, W.-K. and Cohen, E. (2008), Anisotropic geodesic distance computation for parametric surfaces, *in* 'Shape Modeling and Applications, IEEE International Conference on', SMI '08, pp. 179–186.

Sethian, J. A. (1995), A fast marching level set method for monotonically advancing fronts, *in* 'Proc. Nat. Acad. Sci', pp. 1591–1595.

Shen, X. and Wu, Y. (2010), Sparsity model for robust optical flow estimation at motion discontinuities, *in* 'Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '10, pp. 2456–2463.

Simpson, I., Cardoso, M., Cash, D., Modat, M., Ourselin, S., Woolrich, M. and Schnabel, J. (2013), A bayesian approach for spatially adaptive regularisation in non-rigid registration., *in* 'International Conference on Medical Image Computing and Computer-Assisted Intervention', Vol. 16 of *MICCAI '13*, pp. 10–18.

Soille, P. (2003), *Morphological Image Analysis: Principles and Applications*, 2 edn, Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Sotiras, A., Davatzikos, C. and Paragios, N. (2013), 'Deformable medical image registration: A survey', *IEEE Trans. Med. Imaging* **32**(7), 1153–1190.

Srikrishnan, V. and Chaudhuri, S. (2011), 'Adaptive smoothness based robust active contours', *Image Vision Comput.* **29**(5), 317–328.

Steidl, G. and Weickert, J. (2002), Relations between soft wavelet shrinkage and total variation denoising, *in* 'DAGM-Symposium', pp. 198–205.

Steinbrucker, F., Pock, T. and Cremers, D. (2009), Large displacement optical flow computation without warping, *in* 'Proceedings of the Twelfth International Conference on Computer Vision', ICCV '09, pp. 1609–1614.

Sun, D., Roth, S. and Black, M. J. (2010), Secrets of optical flow estimation and their principles, *in* 'Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '10, pp. 2432–2439.

Sun, D., Roth, S., Lewis, J. P. and Black, M. J. (2008), Learning optical flow, *in* 'Proceedings of the 10th European Conference on Computer Vision - Volume Part I', ECCV '08, pp. 83–97.

Sun, D., Sudderth, E. B. and Black, M. J. (2010), Layered image motion with explicit occlusions, temporal consistency, and depth ordering, *in* 'Advances in Neural Information Processing Systems', NIPS '10, pp. 2226–2234.

Sun, D., Sudderth, E. B. and Black, M. J. (2012), Layered segmentation and optical flow estimation over time, *in* 'Proceedings of the 2012 Conference on Computer Vision and Pattern Recognition', CVPR '12, pp. 1768–1775.

Sun, D., Wulff, J., Sudderth, E. B., Pfister, H. and Black, M. J. (2013), A fully-connected layered model of foreground and background flow, *in* 'Proceedings of the 2013 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '13, pp. 2451–2458.

Thirion, J. P. (1998), 'Image matching as a diffusion process: an analogy with maxwell's demons', *Medical Image Analysis* **2**(3), 243–260.

Thirusittampalam, K., Hossain, M., Ghita, O. and Whelan, P. (2013), 'A novel framework for cellular tracking and mitosis detection in dense phase contrast microscopy images', *Biomedical and Health Informatics, IEEE Journal of* **17**(3), 642–653.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society - Series B* **58**, 267–288.

Trobin, W., Pock, T., Cremers, D. and Bischof, H. (2008), An unbiased second-order prior for high-accuracy motion estimation, *in* 'Proceedings of the 30th DAGM Symposium on Pattern Recognition', pp. 396–405.

Tylavsky, D. and Sohie, G. R. L. (1986), 'Generalization of the matrix inversion lemma', *Proceedings of the IEEE* **74**(7), 1050–1052.

Unger, M., Werlberger, M., Pock, T. and Bischof, H. (2012), Joint motion estimation and segmentation of complex scenes with label costs and occlusion modeling, *in* 'Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '12, pp. 1878–1885.

Vincent, L. and Soille, P. (1991), 'Watersheds in digital spaces: an efficient algorithm based on immersion simulations', *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(6), 583–598.

Wahba, G. (1977), 'Practical approximate solutions to linear operator equations when the data are noisy', *SIAM Journal on Numerical Analysis* **14**(4), 651–667.

Wainwright, M. J. and Simoncelli, E. P. (2000), Scale mixtures of Gaussians and the statistics of natural images, *in* S. A. Solla, T. K. Leen and K.-R. Müller, eds, 'Advances in Neural Information Processing Systems', Vol. 12 of *NIPS '00*, MIT Press, Cambridge, MA, pp. 855–861.

Wang, T., Cheng, I. and Basu, A. (2009), 'Fluid vector flow and applications in brain tumor segmentation', *IEEE Trans. Biomed. Engineering* **56**(3), 781–789.

Wedel, A., Cremers, D., Pock, T. and Bischof, H. (2009), Structure- and motion-adaptive regularization for high accuracy optic flow, *in* 'Proceedings of the International Conference on Computer Vision', ICCV '09, pp. 1663–1668.

Weickert, J. and Schnörr, C. (2001), 'Variational optic flow computation with a spatio-temporal smoothness constraint', *J. Math. Imaging Vis.* **14**(3), 245–255.

Weiss, Y. (1997), Smoothness in layers: Motion segmentation using nonparametric mixture estimation., *in* 'Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition', CVPR '97, IEEE Computer Society, Washington, DC, USA, pp. 520–526.

Werlberger, M., Pock, T. and Bischof, H. (2010), Motion estimation with nonlocal total variation regularization, *in* 'Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', CVPR '10, pp. 2464–2471.

Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D. and Bischof, H. (2009), Anisotropic huber-l1 optical flow, *in* 'Proceedings of the British Machine Vision Conference', BMVC '09, London, UK.

Wohlberg, B. and Rodriguez, P. (2007), 'An iteratively reweighted norm algorithm for minimization of total variation functionals', *Signal Processing Letters, IEEE* **14**(12), 948–951.

Woodford, O., Torr, P., Reid, I. and Fitzgibbon, A. (2009), 'Global stereo reconstruction under second-order smoothness priors', *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(12), 2115–2128.

Xiao, J., Cheng, H., Sawhney, H., Rao, C. and Isnardi, M. (2006), Bilateral filtering-based optical flow estimation with occlusion detection, *in* 'Proceedings of the 9th European conference on Computer Vision - Volume Part I', ECCV '06, pp. 211–224.

Xie, X. and Mirmehdi, M. (2004), 'Rags: Region-aided geometric snake', *Trans. Img. Proc.* **13**(5), 640–652.

Xu, C. and Prince, J. L. (1998*a*), 'Generalized gradient vector flow external forces for active contours', *Signal Process.* **71**(2), 131–139.

Xu, C. and Prince, J. L. (1998*b*), 'Snakes, shapes, and gradient vector flow', *IEEE Trans. Img. Proc.* **7**(3), 359–369.

Xu, L., Chen, J. and Jia, J. (2008), A segmentation based variational model for accurate optical flow estimation, *in* 'Proceedings of the 10th European Conference on Computer Vision: Part I', ECCV '08, pp. 671–684.

Xu, L., Jia, J. and Matsushita, Y. (2012), 'Motion detail preserving optical flow estimation', *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(9), 1744–1757.

Yang, A. Y., Zhou, Z., Balasubramanian, A. G., Sastry, S. S. and Ma, Y. (2013), 'Fast l1-minimization algorithms for robust face recognition', *IEEE Trans. Img. Proc.* **22**(8), 3234–3246.

Yang, S., Kohler, D., Teller, K., Cremer, T., Le Baccon, P., Heard, E., Eils, R. and Rohr, K. (2008), 'Nonrigid registration of 3-d multichannel microscopy images of cell nuclei', *IEEE Trans. Img. Proc.* **17**(4), 493–499.

Yang, X., Li, H. and Zhou, X. (2006), 'Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy', *Circuits and Systems I: Regular Papers, IEEE Transactions on* **53**(11), 2405–2414.

Yaroslavsky, L. P. and Yaroslavskij, L. (1985), 'Digital picture processing. an introduction.', *Digital picture processing* **1**.

Yuan, J., Schnörr, C. and Steidl, G. (2009), Total variation based piecewise affine regularization, *in* 'Scale Space and Variational Methods in Computer Vision', Vol. 5567 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 552–564.

Yuan, J., Schörr, C. and Steidl, G. (2007), 'Simultaneous higher-order optical flow estimation and decomposition', *SIAM J. Sci. Comput.* **29**(6), 2283–2304.

Zach, C., Pock, T. and Bischof, H. (2007), A duality based approach for re-altime tv-l1 optical flow, *in* 'Proceedings of the 29th DAGM Conference on Pattern Recognition', pp. 214–223.

Zimmer, C., Labruyere, E., Meas-Yedid, V., Guillen, N. and Olivo-Marin, J.-C. (2002), 'Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: A tool for cell-based drug testing', *IEEE Trans. Med. Imaging* **21**(10), 1212–1221.

Zimmer, H., Bruhn, A. and Weickert, J. (2011), 'Optic Flow in Harmony', *Int. J. Comput. Vision* **93**, 368–388.

Zitnick, C. L., Kang, S. B., Uyttendaele, M., Winder, S. A. J. and Szeliski, R. (2004), 'High-quality video view interpolation using a layered representation.', *ACM Trans. Graph.* **23**(3), 600–608.