Technical University of Crete, Greece
School of Electrical and Computer Engineering

Motion Structure Analysis in Rivers for Evaluation of Dangerous Events

Konstantinos Bacharidis

Thesis Committee
Professor Michail Zervakis (ECE)
Associate Professor Michail G. Lagoudakis (ECE)
Associate Professor Ioannis Papaeustathiou (ECE)

Chania, October 2016

Πολυτεχνειο Κρητης

Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων

Ανάλυση της Δομής Κίνησης σε Ποτάμια για την Αξιολόγηση Επικίνδυνων Συμβάντων

Κωνσταντίνος Μπαχαρίδης

Εξεταστική Επιτροπή

Καθηγητής Μιχάλης Ζερβάκης (ΗΜΜΥ)

Αναπληρωτής Καθηγητής Μιχαήλ Γ. Λαγουδάκης (ΗΜΜΥ)

Αναπληρωτής Καθηγητής Ιωάννης Παπαευσταθίου (ΗΜΜΥ)

Χανιά, Οκτώβριος 2016

Abstract

Motion estimation is a widely researched topic of Computer Vision with numerous published algorithms. Nowadays, Computer Vision and motion estimation are utilized to confront real world problems. One such application, with on-going research interest, is the estimation of the motion field of rivers using video data. The incorporation of Computer Vision in river flow estimation can lead to the development of a low-cost, fast, accurate and above all, a non-intrusive method of measuring the river's velocity. These characteristics are of most importance since until now, this estimation required on-field measurements with costly conventional equipment, such as accelerometers or doppler-based devices.

Moreover, natural phenomena that have rapid occurrence, such as flash flood streams, are hard to monitor at their full extent with the use of conventional equipment. On the other hand, a video surveillance system accompanied with a motion estimation algorithm would allow constant and accurate measurements of the flow without requiring the on-field presence of hydrologists or any other man-held equipment, under the condition that a light-source is available. However, the development of a motion estimation algorithm for the case of fluids can be a difficult task due to the dynamic motion nature of the fluid, guided by both internal and external forces, such as wind and gravity. In addition, when applying such methodologies to real world problems, such as the development of a river monitoring system, researchers must also deal with a number of other tasks beyond the selection of an appropriate motion estimation algorithm. These tasks span from the selection of hardware equipment to be used and the monitoring layout formation as well as the data mining capabilities that offers, to even the derivation of a way to associate the 2-dimensional image space with the 3-dimensional real world in order to reach to a real world velocity estimate for the river's velocity.

This thesis addresses the problem of developing an Image-based monitoring layout for river flows presenting and discussing the methodologies behind each module. More specifically, we propose the utilization of a stereo camera monitoring layout allowing the relation of the real world and image plane coordinate systems and thus the 2-Dimensional image plane motion and the 3-Dimensional river surface motion. Moreover, a novel probabilistic motion estimation method based on stereoscopic data is presented which evaluates each possible displacement based on the two views of the monitored scene resulting in a dense global motion field that accurately represents the motion patterns present on the

river's surface. However, this thesis does not comments upon the performance and cost trade-off of the used hardware.

Περίληψη

Η εκτίμηση κίνησης αποτελεί ένα ευρέως εξεταζόμενο θέμα στον επιστημονικό κλάδο της Μηχανικής Όρασης έχοντας μάλιστα πληθώρα δημοσιευμένων αλγορίθμων και μεθοδολογιών. Στις μέρες μας, η Μηχανική Όραση και η εκτίμηση κίνησης χρησιμοποιούνται για την αντιμετώπιση προβλημάτων στον πραγματικό κόσμο. Μια τέτοια εφαρμογή, με εξελισσόμενο επιστημονικό ενδιαφέρον, αποτελεί η εκτίμηση του πεδίου κίνησης ποταμών αξιοποιώντας δεδομένα βίντεο. Η ενσωμάτωση του κλάδου της Μηχανικής Ορασης στην εκτίμηση του πεδίου κίνησης των ποταμών μπορεί να οδηγήσει στην ανάπτυξη μιας χαμηλού κόστους, γρήγορης, ακριβέστατης και πρωτίστως μη επεμβατικής μεθόδου εκτίμησης της επιφανειακής ταχύτητας ενός ποταμού. Αυτά τα χαρακτηριστικά τυγχάνουν ιδιαίτερης σημασίας δεδομένου ότι μέχρι τώρα, η εκτίμηση της ταχύτητας ενός ποταμού απαιτούσε μετρήσεις πεδίου χρησιμοποιώντας δαπανηρό συμβατικό εξοπλισμό, όπως επιταχυνσιόμετρα ή συσκευές βασιζόμενες σε τεχνολογία Doppler.

Επιπροσθέτως, φυσικά φαινόμενα που έχουν ταχεία εμφάνιση, όπως πλημμυρικά φαινόμενα, είναι δύσκολο να παρακολουθηθούν, σε όλη τους την έκταση με τη χρήση συμβατικού εξοπλισμού. Από την άλλη πλευρά, ένα σύστημα παρακολούθησης βίντεο το οποίο συνοδεύεται με έναν αλγόριθμο εκτίμησης κίνησης μπορεί να επιτρέψει συνεχείς και ακριβείς μετρήσεις της ροής χωρίς να απαιτείται η επιτόπια παρουσία υδρολόγων και η χρήση εξοπλισμού που απαιτεί ανθρώπινο χειρισμό, υπο την προϋπόθεση υπάρξης πηγής φωτός που θα επιτρέπει την απεικόνιση της σκηνής απο την κάμερα. Ωστόσο, η ανάπτυξη ενός αλγορίθμου εκτίμησης κίνησης για την περίπτωση ενός ρευστού μπορεί να αποτελέσει ένα δύσκολο έργο, εξαιτίας της δυναμικής φύσης της κίνησης του ρευστού, η οποία καθοδηγείται τόσο απο εσωτερικές όσο και απο εξωτερικές δυνάμεις, όπως ο αέρας και η βαρύτητα. Επιπλέον, κατά την εφαρμογή μεθοδολογιών βασιζόμενων σε επεξεργασία εικόνας σε προβλήματα του πραγματικού κόσμου, όπως είναι και η ανάπτυξη ενός συστήματος παρακολούθησης ποταμών, οι ερευνητές θα πρέπει επίσης να ασχοληθούν με μια σειρά από άλλα καθήκοντα, πέρα απο την επιλογή του κατάλληλου αλγόριθμου εκτίμησης κίνησης. Τα καθήκοντα αυτά εκτείνονται από την επιλογή του εξοπλισμού που θα χρησιμοποιηθεί και την τοπολογική αξιολόγηση της διάταξη παρακολούθησης σε συνδυασμό με τις δυνατότητες εξόρυξης δεδομένων που προσφέρονται, μέχρι και την εύρεση ενός τρόπου διασύνδεσης του 2-διάστατου χώρου της εικόνας με τον 3-διαστάσεων πραγματικό κόσμο προκειμένου να φτάσει σε μια εκτίμηση της πραγματικής ταχύτητας του ποταμού.

Η παρούσα διατριβή ασχολείται με το πρόβλημα της ανάπτυξης μιας διάταξης παρακολούθησης ροών ποταμών βασιζόμενη σε δεδομένα βίντεο, παρουσιάζοντας και συζητώντας τις μεθοδολογίες που μπορούν να αξιοποιηθούν καθώς το πως εν τέλει αναπτύχθηκαν σε

κάθε ενότητα του συστήματος παρακολούθησης, εξαιρώντας όμως, το σχολιασμό και την αξιολόγηση της σχέσης απόδοσης - κόστους για τον εξοπλισμό που χρησιμοποιήθηκε στο πρωτότυπο του συστήματος. Συγκεκριμένα, προτείνουμε την χρήση μιας στερεοσκοπικής διάταξης παρακολούθησης η οποία επιτρέπει την συσχέτιση των διανυσματικών χώρων του πραγματικού κόσμου και του επιπέδου της εικόνας, επιτρέποντας έτσι τη διασύνδεση της 2-Διάστατης ταχύτητας στο επίπεδο της εικόνας και της 3-Διάστατης ταχύτητας του πραγματικού χώρου. Επιπλέον, παρουσιάζεται μια νέα πιθανοτική μέθοδος εκτίμησης κίνησης βασιζόμενη στα στερεοσκοπικά δεδομένα η οποία αξιολογεί κάθε πιθανή μετατόπιση με βάση τις δύο προβολές της παρακολουθούμενης σκηνής και η οποία εξάγει ένα πυκνό πεδίο κίνησης που αντιπροσωπεύει με ακρίβεια τα μοτίβα κίνησης που υπάρχουν στην επιφάνεια του ποταμού.

Acknowledgements

First, I would like to thank my advisor Michail Zervakis for his guidance during the course of this thesis.

Next, I would like to thank the members of the DISPLAY lab, namely Dina Moirogiwrgou and George Livanos who supported me throughout the developments and experimental testing of the system

My friends from Chania Nikos(Kariol..), Bill(Pous..), Christos(Eazer), Vasilis(Pez) and Vasilis(Psilos) with whom I shared some amazing moments during the last three years. My girlfriend Dimitra(D) for her support and encouragement through the writing of this thesis. My close friends from my home town and my University.

Last, but not least, I would like to thank my family for their love, support, and constant encouragement.

Contents

1 Introduction **1**

1.1 Thesis Focus and Contribution 1

1.2 Thesis Outline 3

2 Problem Set-up **5**

2.1 Image-based River Monitoring 5

2.2 Related Work and Component Analysis of Image-based River Monitoring Systems 6

2.2.1 Flow Visualization and Recording 6

2.2.2 Image-based Fluid Motion Estimation 8

2.2.3 2D to 3D Motion field Transition 12

3 Theoretical Background on Motion Estimation and Stereo Vision **15**

3.1 Image-based Motion Estimation in Fluids 16

3.2 Particle-based Methods 17

3.2.1 Particle characteristics and seeding 18

3.2.2 Particle tracking and motion estimation 21

3.2.3 Particle Image Velocimetry (PIV) 23

3.2.4 Particle Tracking Velocimetry (PTV) 28

3.3 Probabilistic Methods 31

3.3.1 Conditional Probability Models 31

3.3.2 Prior Probability Models 33

3.3.3 Optical Flow Estimation 33

3.4 Differential Methods based on Fluid Properties 34

3.4.1 Methods based on Fluid Mechanics 35

3.4.2 Methods based on the Physical Properties of Fluid Phenomena . . 36

3.5 Stereo Vision . . . 38
3.5.1 Projective Geometry and Estimation . . . 38
3.5.2 Epipolar Geometry . . . 41
3.5.3 Intrinsic and Extrinsic Parameters . . . 42
3.5.4 Radial and Tangential Lens Distortion . . . 45
3.5.5 Camera Calibration . . . 47
3.5.6 Stereo Alignments and Image Plane Relation . . . 52
3.5.7 3D Structure Reconstruction via Triangulation . . . 57

4 Our Approach **65**
4.1 Stereo Monitoring Layout . . . 66
4.1.1 Physical to Image plane coordinates . . . 68
4.1.2 Image Plane relation in the camera pair . . . 69
4.1.3 Disparity and Depth map estimation . . . 72
4.2 Optical flow Estimation . . . 73
4.2.1 Probabilistic Optical flow Formulation . . . 74
4.2.2 Stereoscopic Data Utilization . . . 77
4.3 Velocity Computation in World Coordinate System . . . 81
4.3.1 River Isolation and Main trend Extraction . . . 81
4.3.2 River Surface Velocity Computation . . . 83
4.4 Velocity Estimate Validation . . . 84
4.4.1 Validation Process . . . 85
4.4.2 Region Formulation . . . 87

5 Results **89**
5.1 Improvement of 2D motion estimation using stereo data . . . 90
5.1.1 Accuracy using Real Data . . . 90
5.1.2 Accuracy using Synthetic Dataset . . . 91
5.2 Efficiency of Motion Estimation in Capturing Local trends . . . 95
5.3 Summarization and Clustering of Motion Directions . . . 103
5.4 River Motion Estimation in World Coordinates . . . 107
5.5 Visual Environment: Fluid Flow Viewer . . . 116

6 Conclusion **119**
6.1 Conclusion . . . 119

6.2 Future Work 120
6.2.1 Proposed Image-based Monitoring System 120
6.2.2 Stereo-based Probabilistic Optical flow Method 120

References **129**

List of Figures

2.1 Monitoring layout deployment example of existing image-based river monitoring systems. The selected area is videotaped by a camera laid on top of a bridge. 7

2.2 Cross-correlation particle tracking process applied in a small cluster of particles. Image taken from Muste et.al. [48]. 9

2.3 Positioning and relation between the camera and the ground truth points. Figure taken from Muste et.al. [48]. 12

2.4 Orthorectification application to a river view. Image taken from Muste et.al. [61]. 13

3.1 A laser based monitoring procedure, which uses laser beams as tracer pointers. Such monitoring layouts are preferable for small scaled flows where PIV motion estimation schemes can be used. 19

3.2 An example of a window deformation approach in a two-stage search with the interrogation window and the search area being deformable. At the 1st stage we have a square shaped window whereas in the 2nd stage becomes a circle around the best matched pixel found in the 1st stage. 23

3.3 Generalized Cross-Correlation scheme using the Phase Only Filter $W(i, j)$ defined by the spectral transformation of the scene magnitude. 25

3.4 The Translational Stereoscopic camera alignment. 27

3.5 The Rotational Stereoscopic camera alignment. 28

3.6 Relation between the 3-D world, the 3-D camera, and the 2-D image plane coordinate systems. 39

3.7 (a)Image with perspective distortion with the windows not being rectangular and their lines converging at a finite point, (b)Image taken from Frank-furt airport. Line parallelismis lost with lines converging also to finite point. 40

3.8 (a)Point Correspondence, with C and C' being the camera centres of the image planes. The 3D space point X, and the image plane point correspondences images ($xandx'$) lie in the common epipolar plane π. (b) The epipolar line define a ray in which the 3D space point X must lie in order the image of the X to lie on the epipolar line l' in the the second view. Figure was taken from Hartley and Zisserman [42]. 42

3.9 Extrinsic parameters. 43

3.10 Radialdistortion variations. 46

3.11 Tangential distortion, ... 47

3.12 (a)Parallel stereo rig, (b) Convergent stereo rig. 53

3.13 The rectification process of transforming the image pair, with C, C' being the principal points, and [R|t] the transformations relating the two planes. 54

3.14 (a) Chessboard view from a convergent stereo rig prior rectification with disparities present in both x and y axis, (b) Rectified Chessboard view with disparities only present in the x-axis. 57

3.15 The reconstruction transformations for the case of the building view. Image transformed based on an image presented in Hartley and Zisserman [42]. 62

3.16 The The projective reconstruction moved to metric through the use of five (or more) world points: (a) the five points placed on the initial scene and, (b) the reconstruction after the point mapping. Image taken from Hartley and Zisserman [42]. 63

4.1 (a)Non Convergent Stereo Rig, (b) Convergent Stereo Rig, placed on a bridge. 67

4.2 (a) The distorted scene representation as recorded from the camera, (b) the undistorted scene representation after the orthorectification process. . 71

4.3 The relation of corresponding points in the two views after the plane rectification process. 73

4.4 Relation between the stereo image pairs and their optical flow estimates based on the disparity map. 78

4.5 River isolation via segmentation, (a) Initial image of Koiliaris river, (b) Class containing vegetation, rock formations, leaves and (c) Class containing the pixels belonging to the river. 82

4.6 Formulation of the search area for the particle in each image plane by defining a window with a horizontal size equal to the image's width and vertical size the size of the particle ± M pixels up and down. The window is formulated around the centroid of the second indicator. 87

5.1 (a)The estimated motion field using a single camera and stereo camera layout,and, (b)Estimated motion field using the on-field stereo layout scheme. 91

5.2 (a) The synthetic reconstructed river scene with motion added on the leaves, (green) indicates the initial position and (magenta) the shifted position (a translational motion to the right direction), (b) the corresponding ground truth motion field, (c) the estimated optical flow field using the single camera probabilistic optical flow method, and (d) the estimated optical flow field using the stereo camera probabilistic optical flow method. . . . 92

5.3 Average Angular and Endpoint Errors and their standard deviations using various window sizes in the examination neighborhood formulation for the case of the stereo-based probabilistic optical flow estimation method. (a) AE and StD of AE for both methods, (b) EE and StD of EE for both methods, (c)AE, EE for the proposed stereo-based approach compared to the optimal neighborhood window size found for the single camera probabilistic optical flow estimation method(13 x 13 window- marked with blue, bullet points indicate the corresponding AE and EE). 94

5.4 Linear shift case; (a) Ground Truth motion field, (b) Stereo-Based Probabilistic method(**Proposed**), (c) Single Camera Probabilistic approach, (d) Weighted Lucas - Kanade, (e) Horn-Schunck and (f) OpenPIV tool. . . . 97

5.5 Vortex pair case; (a)Ground truth motion field, (b) Stereo-Based Probabilistic method(**Proposed**), (c) Single Camera Probabilistic approach, (d) Weighted Lucas - Kanade, (e) Horn b•“ Schunck and (f) OpenPIV tool. . 99

5.6 Membrane case; (a) Ground truth motion field, (b) Stereo-Based Probabilistic method(**Proposed**), (c) Single Camera Probabilistic approach, (d) Weighted Lucas - Kanade, (e) Horn-Schunck and (f) OpenPIV tool. . . . 102

5.7 (a) Initial frame(red box denotes the initial dimensionally known work area), (b) Initial work area(white box) based on Q-liner's length dimension, (c) Expanded work area, (d) Optical flow field of the work area, and, (e) Classification result of the underlying motions 104

5.8 Main trend of motion classification deviation between K-NN and Naive Bayes classifiers prior the global optimization step, (a) K-NN with 4 classes, real-world velocity estimate=0.6350 m/sec, (b) Naive Bayes classifier with 4 classes, real-world velocity estimate=0.6377 m/sec, (c) K-NN with 5 classes, real-world velocity=0.7820 m/sec and (d) Naive Bayes with 5 classes, real-world velocity estimate=0.7334 m/sec. 105

5.9 Koiliaris River site,(a)The convergent stereo rig placement on top of the bridge, (b)Close up on the stereo rig used, consisting of two CCD cameras, (c)View of the Left camera, (d)View of the Right camera without rectification. 108

5.10 An example of the estimated motion fields and the derived main trends using stereo camera layout and stereo-based probabilistic optical flow method. 109

5.11 Koiliaris river test case, (a) river's motion field without projection correction placed in image coordinate system and (b) scene and motion field rescaled motion field based on the reconstructed 3D scene coordinates. . 110

5.12 Leaf bunch tracking in Koiliaris river test case, (a) leaf bunch being tracked and corresponding motion field without projection correction placed in image coordinate system and (b) scene and motion field orthorectification based on the reconstructed 3D scene coordinates. 111

5.13 Particle isolation and tracking, (a) Segmented scene containing the vegetation and the natural particle being tracked, i.e. the leaf, (b) Isolated particle for the examination process, (c) The vegetation and natural particles present at N = 10 frames forward, (d) Isolated particles present in the current frame and (e) Matching of the reference particle in the current frame to verify that it has cascaded to this expected region at this current frame based on the velocity estimate, with red is the reference particle pattern and blue the particles present in the current frame. 114
5.14 Validation stage in Koiliaris River. Red and blue colors indicate the motion between consequent frames. (a) The leaf bunch formation across the frame series are fused into one unique frame to better represent the methodology stages across the frame series, (b) Identified and tracked leaf formation based on a reference leaf pattern taken from previous frames. 115
5.15 Initialization window of Fluid Flow Viewer (v.1.0). 116
5.16 Some of the Stages of FFV, (a)Scene view and Frame capture, (b)Camera Calibration and Error Visualization,(c)Depth Map Creation(*image from another session*, (d)Indicator Selection, (e) Region formulation and Motion field 118

Chapter 1

Introduction

Computer Vision is a diverse and active science field with its researchers showing various interests and aspirations to develop methodologies that confront real world problems, and thus, blending Computer Vision with other scientific fields. One such attempt is to combine Computer Vision with Hydrology, developing image based estimation and inference methods concerning the fluid's motion field.

An application of this scientific marriage is in river and stream flow monitoring. The estimation of the flow field of a river or a stream is performed traditionally with the use of conventional equipment, such as accelerometers, which require on-field measurements. Such monitoring methods, despite the costly human operated equipment do not provide continuous monitoring of the flow. This aspect is crucial, in cases of natural hydrological phenomena that have rapid outbreaks, such as flash floods. Hydrologists rarely can monitor cases of such phenomena at their full extent since no warning about the imminent occurrence is possible. In addition, the majority of the conventional equipment that is used requires the presence of hydrologists near or even within the flow region of the stream or river. This means that in extreme cases of flash flood events with rapid flow increase, the measurement acquisition process can become dangerous for the researchers.

1.1 Thesis Focus and Contribution

This thesis describes a new image-based river and stream flow monitoring framework. The monitoring module integrates image data and depth information of the monitored

scene, captured with the use of a stereo camera layout, estimating the river's average real-world surface velocity at any time.

The river's flow field is initially estimated in the image plane domain, with the use of a probabilistic optical flow estimation methodology that combines the image data from the two views to extract displacement probabilities for each examined image region. The prediction of the position of each object in the next time frame follows a Bayesian inference scheme based on a number of assumptions about the relation between neighbouring regions as well as the motion of each pixel within the examined region leading to dense optical flow field estimates. In essence a stochastic motion estimation model is applied for the derivation of the motion field making this method ideal for the estimation of fluid's flow field since its dynamic motion nature results in the relative positions of neighbouring points to change faster in a fluid than on a rigid body.

This methodology, initially presented by Chang et.al. [1] based on a 3-Dimensional (time and space) neighbourhood, was redesigned in my undergraduate diploma thesis (Bacharidis [2]) to allow a 2-Dimensional neighbourhood refining prediction ambiguities due to the data reduction. In this thesis, we introduce an new theoretical basis on the estimation and neighbourhood formulation process by incorporating the image data from the two scene views acquired by the stereo camera layout. This new estimation scheme achieved the same or in some cases increased estimation accuracy compared to the previous approach, requiring at the same time approximately 16% less amount of data.

Moreover, to exclude disambiguous and erroneous motions within the flow produced due to diffusion effects by occluding rocks with the river or at the river banks we applied clustering and classification techniques on the estimated motion field in order to extract the main motion trend of the flow. A variety of clustering and classification Machine Learning techniques were examined, leading to the observation that a supervised classification approach that utilizes a prior knowledge about the expected motion of the river flow can achieve higher classification accuracy, with Naive Bayes yielding the best results.

Finally, in order to associate the 2-Dimensional motion field with the real world, thus, estimating the 3-Dimensional velocity, we utilized the scene's depth information derived using the stereo layout. The relation between the projection planes allowed the association of the position of each object in the 2-Dimensional image plane with its corresponding position in the 3-Dimensional world coordinate system. This association allowed

the relation between the 2-Dimensional pixel displacements and their corresponding 3-Dimensional displacements, generating the 3-Dimensional motion field, and thus allowing the derivation of the 3-Dimensional motion field of the river. In this thesis, we will describe and present the theoretical basis behind these relations as well as the constraints and assumptions that were made to achieve an accurate projection relation between the two coordinate systems.

The framework was incorporated in a newest version of a fluid flow extraction and visualization tool, Fluid Flow Viewer (F.F.V.), which was initially presented in my undergraduate thesis. This new version allows a user to calibrate the stereo camera layout, extract the depth map of viewed scene, view of the viewed flow from each camera, extract the 2-Dimensional motion fields and finally, estimate the 3-Dimensional average river surface velocity.

1.2 Thesis Outline

In Chapter 2 we state the river flow estimation problem with the use of video data and we refer to different image-based river monitoring approaches that have been developed by other researchers, presenting and commenting on the advantages and disadvantages of each method. In Chapter 3 we present all the background information needed for this thesis. We give an overview of fluid flow estimation problem and present the state of the art image based estimation methodologies. Furthermore, we provide basic information about Stereo Vision and projection relations and scene reconstruction methodologies that will be used in the depth map generation process. In Chapter 4 we describe our river monitoring framework, presenting our novel optical flow estimation method, our main motion trend method extraction scheme, and finally, the relation derivation and transformation scheme that allows the extraction of the 3-Dimensional river velocity. In Chapter 5 we evaluate the accuracy of the presented monitoring framework comparing the velocity estimate with estimates acquired with the use of conventional equipment, we also compare the new stereo data based optical flow estimation algorithm with our previous 2-Dimensional single camera version (presented in Bacharidis [2]) as well as with other known pre-existing optical flow estimation method, and we also present the new version of our graphical user interface(GUI), Fluid Flow Viewer(F.F.V.). Finally,

Chapter 6 acts as an epilogue for this thesis, presenting our conclusions along with future improvements.

Chapter 2

Problem Set-up

2.1 Image-based River Monitoring

Traditionally, river flow monitoring was performed using man-held, expensive equipment, such as accelerometers and doppler-based devices. Moreover, such monitoring methods did not provide continuous monitoring data, a crucial factor for the task of observing and analyzing the behaviour and characteristics of rapid duration fluid phenomena, such as flash floods. The development of an image-based river monitoring system provides non-intrusive, low-cost and constant monitoring of the river flow.

An image based river monitoring system depends on many factors that determine the estimation accuracy, ranging from the appropriate selection of the hardware and the motion estimation algorithm, to the camera position in the monitoring scene. Moreover, the transition to real world conditions introduces new challenges towards producing accurate estimates, thus, special care must be given to pre and post processing of the collected data in order to remove distortions introduced by the external monitoring conditions.

There have been a few approaches which try to estimate the river's motion through the development of an image-based monitoring system. However, most of these methods rely upon the existence of artificial and physical tracers within the flow in order to determine the 2-Dimensional image motion field of the river. Furthermore, the transition from the 2-Dimensional optical to the real world 3-Dimensional flow field and finally, to an average real world surface velocity requires knowledge upon the scene (usually use of ground truth points). These characteristics result in these systems not being fully automated and adaptable to any monitoring scene, leading to gaps that our approach

attempts to fill. The following section will attempt to present the basic ideas behind the implementation of the existing image-based river flow monitoring systems.

2.2 Related Work and Component Analysis of Image-based River Monitoring Systems

We will begin by presenting the key parts behind each image-based river monitoring system analyzing in each component the differences and variations that each existing approach offers, commenting at the same time upon the limitations or advantages that each variation presents.

An image-based river monitoring system essentially consists of three major components, (a) Flow Visualization and Recording, (b) Image Processing and Motion Estimation, and (c) 2-Dimensional to 3-Dimensional motion field association. All three components are related to each other, since there is a pipeline of information flowing from one component to another.

2.2.1 Flow Visualization and Recording

The first component deals with the monitoring hardware used, the camera position in accordance to viewed scene and the affect of the position selection in the data acquisition process and quality. In this section we will also, deal with flow visualization process from the scope of tracer selection.

Monitoring layout formation: Existing image-based river monitoring systems use single camera rigs, most commonly high resolution CCD cameras, placed on-top of bridges at an angle that allows a wide view of the river surface as well as the river banks (e.g. Bradley et. al. [46]; Tsubaki et.al. [47]; Muste et. al. [48]), as shown in figure 2.1. The latter viewing condition is essential to the majority of the developed systems since the ground truth points used to relate the optical and real world velocity are placed in the river banks. The selection of the viewing angle is crucial in order to reduce the spatial distortion that the camera lens introduces, usually angles between 5 to 7 degrees are selected.

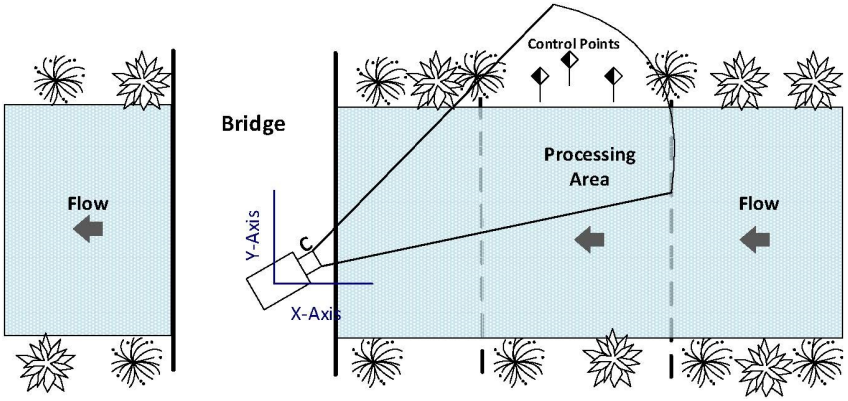


Figure 2.1: Monitoring layout deployment example of existing image-based river monitoring systems. The selected area is videotaped by a camera laid on top of a bridge.

The appropriate selection of the viewing position, when monitoring takes place in real world conditions, is crucial since we must compensate and reduce to best possible ratio the affect of illumination variation of the natural light as well as to be able to acquire the best possible view of the river surface. The illumination variation problem results in glare and shadowing phenomena that can reduce the quality of the recorded image. In order to reduce the image degradation affect due to illumination variation, image pre-processing is applied to the recorded image prior the motion estimation process. The most preferable techniques are, (a) histogram equalization (e.g. Kim [49]) or (b) to deal with uneven illumination as an addictive signal and try to subtracted it from the original image to reveal the detailed information (e.g.Wang et.al. [50];Avgerinakis et.al. [51]). Finally, as far as the recording frame rate of the flow is defined by the lighting conditions and the existence of tracers (for particle-based systems), usually a frame rate of 30 frames per second is selected. The size of the image and the resolution is defined by capability to distinguish movement of the water body in image pairs, as well as the selected particle patterns.

Flow Visualization: The second factor in the first component of an image-based river monitoring system flow visualization tracer selection. The majority of the existing approaches rely on the existence of tracers, natural(e.g. foam or boils) or artificial, to apply particle tracking in order to produce the motion field(e.g. Fujita and Komura [52]; Bradley et. al. [46]; Muste et.al. [48]). The biggest problem in particle-depended systems is insufficient flow seeding. In cases where no natural tracers are present, artificial tracers

need to be added, thus, reducing the system automation and constant monitoring characteristics. Moreover, for the case of artificial tracers as will be mentioned in Chapter 3, careful selection of their characteristics (material, volume e.t.c.) should be made in order to ensure the validity of association of their velocity field with the river's.

Another case of flow visualization, unconstrained by the particle existence condition, is the one when we can use as tracking surrogate the specular reflection formed by incident light interacting with the free-surface deformations as well as waviness generated by wind or large-scale turbulence structures (e.g. Tsubaki et. al. [47];Creutin et.al. [53]). Such image patterns can be easily traced across the river motion and accurately depict the surface velocity despite their short duration(Fujita et.al. [54]).

2.2.2 Image-based Fluid Motion Estimation

The next component of an image-based river monitoring system is the 2-Dimensional optical flow field estimation of the river's flow. In this section we will present the methodologies applied in existing monitoring systems. Furthermore, we also deal with motion field post-processing approaches that aim to reduce the number of erroneous motion vector estimates. Starting from the task of estimation the 2-Dimensional motion field of the river, existing approaches can be group in two categories (a) Particle-based Approaches, and (b) Surface Deformation Approaches, based on the utilization of particles in the tracking process.

2.2.2.1 Particle-based Approaches

The most widely class of methods used in image based monitoring systems is the Particle Image Velocimetry methodology (PIV). The idea behind PIV approaches is that we to use a characteristic pattern (particle) to approximate the motion field of the fluid under the assumption that motion pattern of the particle follows the motion pattern of the fluid, in our case the water flow in the river or stream.

Existing monitoring systems mainly rely upon cross-correlation and feature-based PIV motion estimation methods to derive the motion field (Bradley et.al. [46]; Fujita et. al. [55]; Muste et.al. [48];Fujita and Komura [52]). Cross-correlation and feature-based methods compute similarity indexes, such as Mean Squared Error (MSE) or Sum of Squared Differences (SSD), between an interrogation region in the first image and

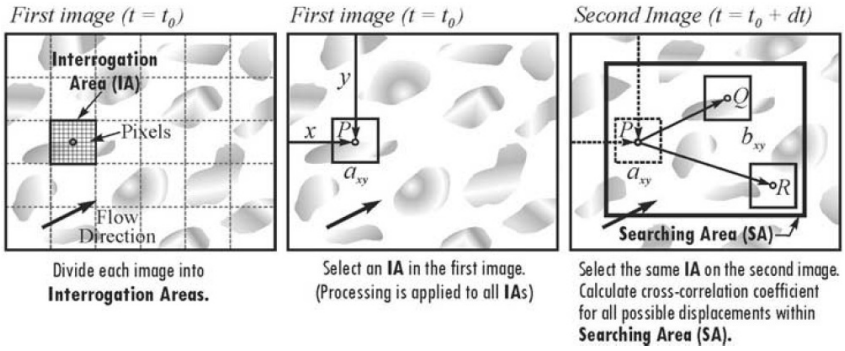


Figure 2.2: Cross-correlation particle tracking process applied in a small cluster of particles. Image taken from Muste et.al. [48].

candidate interrogation regions in a search area in the subsequent frame, as shown in figure 2.2. The interrogation region in the first image can be either applied directly on the tracked particle pattern, initially isolated using color or feature based segmentation(e.g. Harris Matrix) from the river flow, or on the entire image and then isolating the motion vectors belonging to the pattern. For example, Fujita et.al. [55] as well as Bradley et.al. [14], used a cross-correlation coefficient $R_{a,b}$ (Pearson's coefficient) as a similarity index to track a particle pattern across the frames and estimate the displacement vectors:

$$R_{a,b} = \frac{\sum_l \sum_k \left(a_{kl} - \bar{a_{kl}}\right)\left(b_{kl} - \bar{b_{kl}}\right)}{\sqrt{\sum_l \sum_k \left(a_{kl} - \bar{a_{kl}}\right)^2 \sum_l \sum_k \left(b_{kl} - \bar{b_{kl}}\right)^2}} \tag{2.1}$$

where a_{kl} are the grey-scale values for pixels in the interrogation spot, and b_{kl} are the grey-scale values for the corresponding pixels in the search area and $\bar{a_{kl}}, \bar{b_{kl}}$ the average intensity values in the interrogation spot and the search area.

The accuracy of cross-correlation and feature -based approaches depends on the search area selection. The size of the search window should be appropriately selected in order to ensure the in-plane motion of the particle is captured. A fixed location for the interrogation region in the second image results in velocity bias errors and high signal-to-noise ratio due to large displacements (Adrian [56]). In order to solve this problem we can define the interrogation as well as the search area using multi-stage window deformations (see figure 3.2), as will be presented in Chapter 3. Such concepts have been extensively applied in existing image-based river monitoring systems, such as the one presented by Muste et.al. [48], who applied a decoupling approach of the interrogation area to any arbitrary

location in the second image, thus, allowing a dynamic range of velocity measurements and improving the signal-to-noise ratio.

2.2.2.2 Surface Deformation Approaches

As previously mentioned, the second class of motion estimation approaches utilized in image-based river monitoring rely upon river surface deformations due to spectral reflection of the natural light or waviness introduced by external forces, such as wind and gravity (Tsubaki et. al. [47];Creutin et.al. [53]; Fujita et.al. [54]). Such approaches are applied to the entire image and no particle detection and tracking is required since they are not bound by the existence of particles. In order to compute the motion field of the image such methods rely upon the image gradients as a means of luminance distribution propagation along the recorded image frames (Tsubaki and Fujita [57]; Fujita et.al. [54]) or even cross correlation metrics, such as MSE and SSD, searched in a predefined search region (e.g.Creutin et.al. [53]).

A representative method on this concept is Space-Time Image Velocitry(STIV), presented by Tsubaki and Fujita (Tsubaki and Fujita [57]; Fujita et.al. [54]). For image pattern an orientation angle of speed propagation is computed based on image gradients. The reference pattern is compared to candidate patterns in subsequent frames through a coherence measure C computed based on the orientation angle:

$$C = \frac{\sqrt{(J_{xx} - J_{tt})^2 + 4J_{xt}^2}}{J_{xx} + J_t} \tag{2.2}$$

where C is the Coherence measure, J_{xx}, J_{tt}, J_{xt} are structure tensors calculated as follows:

$$J_{xx} = \int_A \frac{\partial g}{\partial x}\frac{\partial g}{\partial x} dxdx \tag{2.3}$$

$$J_{tt} = \int_A \frac{\partial g}{\partial t}\frac{\partial g}{\partial t} dtdt \tag{2.4}$$

$$J_{xt} = \int_A \frac{\partial g}{\partial x}\frac{\partial g}{\partial t} dxdt \tag{2.5}$$

where

$$\frac{\partial g}{\partial x} = \frac{g_{i+2} - 8g_{i+1} + 8g_{i-1} - g_{i-2}}{12\Delta x} \tag{2.6}$$

where $\frac{\partial g}{\partial x}$ is the 4^{th} order central difference scheme and $g(x,t)$ is the gray intensity level.

Finally, in order to compute the velocity the pattern is tracked along an interrogation line through the frame series (Space-Time) and the mean orientation angle ϕ of the pattern is used to compute the velocity of the river:

$$U = \frac{S_x}{S_t} tan\phi \tag{2.7}$$

where U is the average velocity, S_x is the length scale of the pattern in m/pixel, S_t is the is the unit time scale of the time axis in sec/pixel and ϕ is the mean orientation of the pattern along the interrogation line.

The main difference between particle and surface deformation approaches is that the first compute the instantaneous velocity of the river whereas the latter produces the mean velocity of the river surface.

2.2.2.3 Erroneous Motion Vector Exclusion

In order to increase the accuracy of the estimated motion field, image-based monitoring systems must refine the estimated motion fields to exclude erroneous motion estimates produced either by diffusion phenomena near the river banks or low signal-to-noise ratio in the image or insufficient correlated particle images for the case of particle-based approaches. Various approaches have been proposed that can correct erroneous motion estimates (e.g. Fujita and Kaizu [59]; Nogueira et.al. [58]) with the idea behind them being the combination of the physical characteristics of the flow with an assumption of local continuity among the vectors. For example, Nogueira et.al. [58] presented a method in which local coherency of a vector with its 8 neighbors is defined as:

$$coh = \frac{\sum_i |u_i - u_o|}{\sum_i |u_i|} \tag{2.8}$$

with u_i is the velocity vector of the eight neighboring points and u_o the velocity vector of the examined point. This metric of coherency is then used to define a measure of uniformity among neighboring vectors, thus, allowing the correction of non-coherent neighboring vector. Such approaches have been incorporated to existing image-based river monitoring systems, increasing the estimation accuracy (Fujita [55]; Muste et.al. [48]).

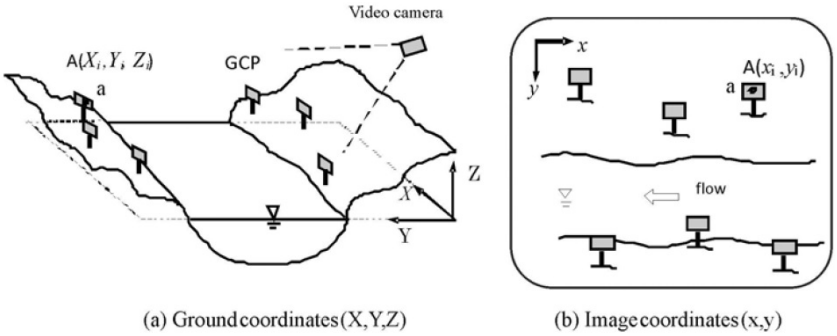


Figure 2.3: Positioning and relation between the camera and the ground truth points. Figure taken from Muste et.al. [48].

2.2.3 2D to 3D Motion field Transition

The last stage an image-based river monitoring system involves the relation of the 2-Dimensional motion field estimate with its corresponding 3-Dimensional real world. In order to do so we need to define a mapping between the image and world coordinate systems. In single camera systems the way to do this is to use ground truth points (GPs), i.e. points whose 3-Dimensional world coordinates are known. These points are usually placed in the river banks facing the camera, as shown in figure 2.3 .

The association between the GPs and their corresponding image plane points is performed through an eight parameter projective transformation, initially presented and used in an image-based river monitoring system by Fujita et.al. [55]. This process known as *image orthorectification*, has been since then the most common image to real world relation method used in the majority of existing monitoring systems (e.g. Bradley et.al. [46]; Fujita et.al. [54]; Muste et.al.([48], [61]); Creutin et.al. [60]). The relation between the two coordinate systems based on the eight parameters is defined as:

$$x_w = \frac{\alpha_1 \cdot x + \alpha_2 \cdot y + \alpha_3}{\alpha_7 \cdot x + \alpha_8 \cdot y + 1} \tag{2.9}$$

$$y_w = \frac{\alpha_4 \cdot x + \alpha_5 \cdot y + \alpha_6}{\alpha_7 \cdot x + \alpha_8 \cdot y + 1} \tag{2.10}$$

with (x_w, y_w) being the x and y axis components of the 3D real world point X_{world}, (x, y) the coordinates of the point in the image plane and a_i the transformation coefficients.

Given the appropriate number of control data points (at least 8 control points), these transformation coefficients can be estimated by solving a system of linear equations. In

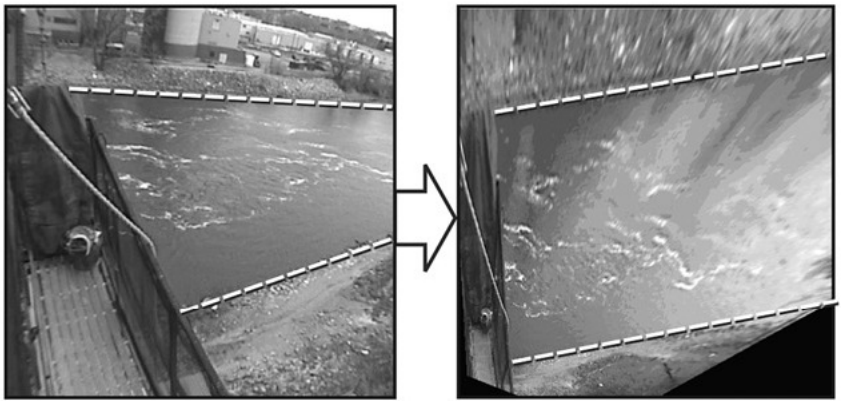


Figure 2.4: Orthorectification application to a river view. Image taken from Muste et.al. [61].

order to apply this approach an horizontal water surface assumption must be made, a condition that also requires the control points to be placed on the horizontal surface. The result of this transformation is an image transformation in which the (x, y) coordinates of the each point is mapped to its corresponding (x_w, y_w) real world coordinates(see figure 2.4). Thus, we can express the pixel displacement into real world displacement based on the aforementioned relation as:

$$\Delta x_w = x_{w,new} - x_{w,old}$$
$$\Delta y_w = y_{w,new} - y_{w,old}$$

and then the real world velocity can be estimated by dividing with the required time interval.

Chapter 3

Theoretical Background on Motion Estimation and Stereo Vision

In this chapter we will further present, analyze and compare the motion estimation approaches that are being used in the 2-D motion field extraction from image data in the existing image-based monitoring systems as well as the overall fluid motion estimation problem. Each method class utilizes different constraints in intensity variation model that leads to the motion field extraction. For example, in particle based methods, an intensity conservation assumption is being made in order to simplify the particle tracking process between frames. On the other hand, some methods utilize fluid mechanics to model, constrain and estimate the intensity variation between the frames. In essence, the difference between the method lies on the constraints applied in the variation of intensity:

$$\nabla I \cdot \boldsymbol{u} + I_t = \textit{Variation Constraint}$$

where $\nabla I = (I_x, I_y)$ is the spatial intensity gradient and $\boldsymbol{u} = (\upsilon, \nu)$ is the optical flow vector. Due to the fact that the time displacement is between two frames,$\Delta t = 1$ the time related factor can be removed.

Moreover, this chapter will provide the theoretical background behind the relation of the 2-D image plane and the 3-D real world using for the case of stereo camera views, as well as the reconstruction ambiguities and constraints introduced based on the deployment characteristics of the stereo layout.

3.1 Image-based Motion Estimation in Fluids

The extraction of the apparent motion of fluids is a difficult task due to the non-rigid motion nature of the fluid. Non-rigid motion is dynamic, meaning that the positions among points in a fluid change constantly and with an unknown scale, due to the effect of external forces, such as wind or gravity, that alter the motion. In Fluid Mechanics the velocity field u constant density and temperature and a pressure field p is described using the Navier-Stokes equations [3] as follows:

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \left(u\dot{\nabla}\right) u = -\frac{1}{\mathrm{p}}\nabla p + v\nabla^2 u + f \tag{3.1}$$

$$\nabla \cdot u = 0 \tag{3.2}$$

with v being the fluid's kinematic viscosity, p its density and f an external force, such as the wind. Naiver-Stokes equations describe the variation of these features based on the variation of the fluid's velocity.

If we observe equation (2.1) we can establish a relation between the fluid's velocity variation with Computer Vision and optical flow estimation. Equation (2.1) resembles the optical flow constraint equation, in which the unknown variable is the velocity vector:

$$\nabla I \cdot \boldsymbol{u} + I_t = 0 \tag{3.3}$$

where $\nabla I = (I_x, I_y)$ is the spatial intensity gradient and $\boldsymbol{u} = (\upsilon, \nu)$ is the optical flow vector. Due to the fact that the time displacement is between two frames,$\Delta t = 1$ the time related factor can be removed.

The optical flow constrain equation allows the motion estimation in image data under the intensity conservation assumption between subsequent frames. This assumption is the basis behind the affine translational models that estimate the motion field of rigid models. However, non-rigidity differs from rigid motion since it is characterized by multi-directional and multi-scalable formulations of the body.

If we relate these two equation based on the velocity being the unknown variable we can formulate models that approximate and constrain non-rigid based intensity variation with physically based feature constraints generated by the properties and quantities of fluids.

There have been numerous image-based motion estimation approaches that try to estimated a fluid's motion. Some of these methods incorporate classical optical flow methods, based on intensity conservation hypothesis, applied in mediums within the flow, such as particle tracers, that provide absent rigidity and whose motion field is assumed to follow and thus, approximate the fluid's. Other methods, are based on stochastic modeling the relative position change of pixels, and the relative intensity variation due to non-rigidity in a fluid image as a random variable with a probability distribution function. Lastly, the most elaborate motion estimation methods utilize motion models combined with physically based feature constraints generated by the properties and quantities of fluids in order to model the multi-directional and multi-scalable formulations encountered.

3.2 Particle-based Methods

The simplest image based motion estimation method class is particle based approaches. These methods apply block matching motion estimation techniques combined with correlation matching and feature tracking on mediums, known as particles or tracers, that flow within or in the surface of the fluid.

The key concept is that the velocity field of these particles can be associated with the velocity field of the fluid under the assumption that the particle's motion follows the same motion pattern as the fluids. Such methodologies although being simple to implement are defined by a number of parameters which ensure that the motion field approximation is as accurate as possible. These parameters relate to the characteristics of the particles present in the flow as well as the measurement requirements during the monitoring of the flow.

As a particle we refer to an element or a group of elements (pattern), natural or artificial, that can be used as a tracer, which can be tracked along the frames and whose properties(size, material properties) allow the association of its displacement vector with the displacement vector of fluid's flow field. In fact it is the combination of particles characteristics and monitoring constraints that separate the developed particle based approaches into two major methodologies, the Particle Image Velocimetry method (PIV) and the Particle Tracking Velocimetry method (PTV). The approaches belonging to these method classes, mainly use correlation and feature based image processing techniques to track particles along the frames and extract the motion field. The main difference

between these two classes is in the number of particles being tracked, with PTV methods tracking a single particle along the frame series making it appropriate for low seeding density flows. On the other hand PIV, tracks groups of particles and thus, is applied in medium and high seeding density flows. This difference defines the yielded velocity characteristics, with PIV allowing the estimation of the Eulerian velocity extracted as the average particle motion in the space based on a group of velocity vectors, whereas PTV, describes the Lagrangian motion in a system as the motion vector is extracted by tracking the displacement path of an individual particle (Cenedese [4]).

3.2.1 Particle characteristics and seeding

In the previous paragraphs we mentioned that the main discrimination criterion between particle based techniques is the particle flow density. The particle seeding density depends on two factors, (a) the particles characteristics and, (b) the monitoring approach that is followed to acquire the particle images.

Starting from the second factor, when monitoring fluid flows there exist two monitoring cases, (1) a confined highly restricted flow within a controlled environment , e.g. a tube, and (2) outdoor monitoring case, e.g. rivers or stream. In the first case, the fluid is placed in a controlled flow environment and is seeded with particles that are illuminated with a light source, such as lasers, and with digital cameras recording the flow (see figure 3.1). The illuminated particles can be identified within the fluid flow, using an image-based correlation scheme. Such monitoring layouts provide images with decreased background noise and particle overlapping (Westerweel et.al. [5]; Meinhart et.al. [6]).

To that respect, it is important that the characteristics of particles and the seeding method are carefully selected such that sufficient light is scattered into the fluid to acquire high-quality images for the estimation process. On the other hand, in outdoor monitoring conditions, when extracting the surface velocity, the estimation accuracy is highly dependable on the particle characteristics since they determine how the exposure to external forces, such as the wind, might affect the velocity relation between the particle and the flow.

But what are the particle's characteristics and how can we select them appropriately in order to increase the velocity estimation accuracy?

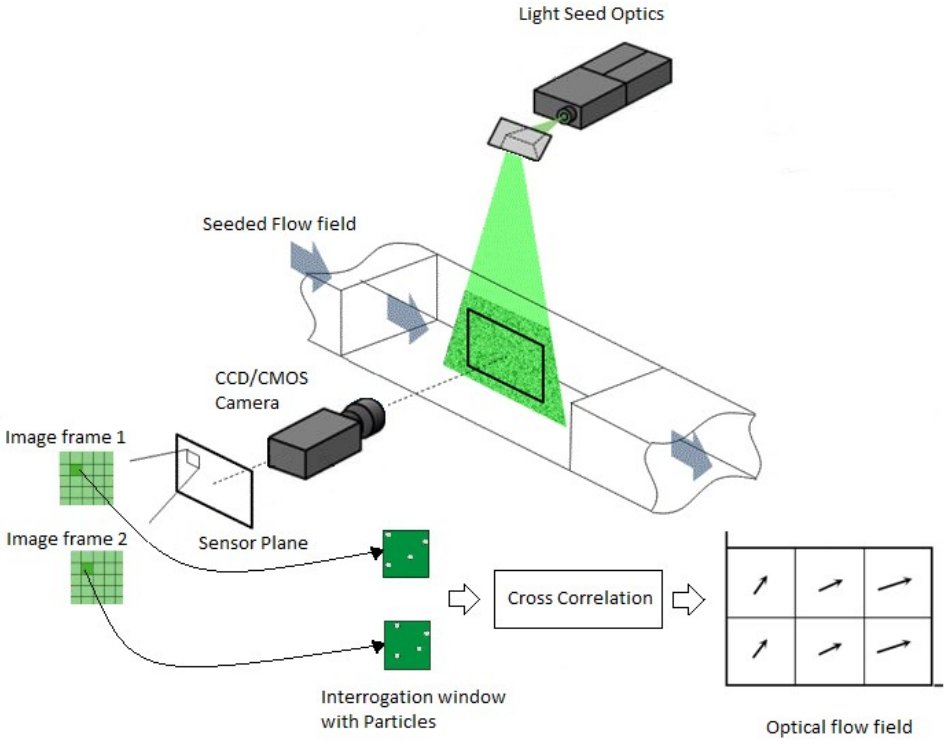


Figure 3.1: A laser based monitoring procedure, which uses laser beams as tracer pointers. Such monitoring layouts are preferable for small scaled flows where PIV motion estimation schemes can be used.

The answer that question is that mainly artificial particles allow to explicitly define appropriately relate their characteristics and seeding density with velocity estimation accuracy. In artificial particles, their characteristics are the diameter, the shape and density. However, thing characteristics are associated also with the fluid's density as well as the dynamic viscosity, meaning that these factors must be taken in mind in order to increase the tracking success of the particle, and essentially the motion estimation accuracy.

Answering to the second part of the question, the factors that determine the appropriate selection particle characteristics, are the particle motion $u_{particles}$, the flow motion u_{flow} as well as the instantaneous relative velocity V_{inst} between the fluid and the particle. An appropriate relation among these factors will allow us to estimate the flow velocity from the particles velocity, based on the particle characteristics. This relation, for the case of spherical artificial tracers is given by the Basset - Boussinesq - Oseen

(BBO) equation which describes the unsteady motion of a suspended sphere based on the aforementioned factors as follows:

$$\frac{\pi d_{particle}^3}{6}\rho_{flow}\frac{du_{particle}}{dt} = \underbrace{-3\pi\mu d_{particle}V_{inst}}_{term1} + \underbrace{\frac{\pi d_{particle}^3}{6}\rho_{flow}\frac{du_{flow}}{dt}}_{term2}\underbrace{-\frac{1}{2}\frac{\pi d_{particle}^3}{6}\rho_{flow}\frac{dV_{inst}}{dt}}_{term3} \underbrace{-\frac{3}{2}d_{particle}^2\left(\pi\mu\rho_{flow}\right)^{1/2}\int_{t_0}^{t}\frac{dV_{inst}}{d\xi}\frac{d\xi}{(t-\xi)^{1/2}}}_{term4} + \underbrace{\sum_k F_k}_{term5} \tag{3.4}$$

with ρ_{flow} being the fluid density, μ the fluid's dynamic viscosity, V_{inst} defining the difference between $u_{particle}$ and u_{flow} , $d_{particle}$ being the particle's diameter, and finally, F_k the existing external forces, for example, gravity.

The basis of the BBO equation is Newton's second law, which in the BBO case describes the imposed forces on the particle's motion:

$$m_{particle}\frac{du_{particle}}{dt} = F_{applied} \tag{3.5}$$

with $m_{particle}$ being the mass of the spherical particle which is equal to:

$$m_{particle} = \frac{\pi}{6}d_{particle}^3\rho_{flow} \tag{3.6}$$

If we combine these two equations we end-up with the BBO equation. Commenting now on the forces applied to the particle (right side of the BBO) these are essentially (a)the viscosity applied to the spherical particle, as defined by Strokes's drag law (*term 1*), (b) the pressure gradient describing the rate as well as the direction in which the pressure is applied and change in the particle space(*term 2*), (c) the mass increase due to the resistance produced to the particle's motion from the fluid's volume (*term 3*), (d) the Basset force, which describes the temporal delay in the boundary layer development due to the relative velocity change over time (*term 4*), and (e) subsequent external forces, such as gravity or wind, etc. (*term 5*).

Essentially, BBO equation provides a correlation between the particle's characteristics, it's velocity and the velocity of the fluid. Thus, an appropriate selection of the particle's characteristics will define the accuracy of the fluid's velocity estimate.

But how do we appropriate select the particle's characteristics?

The condition that ensures the appropriate selection of particle characteristics is the satisfaction of Stoke's drag law. According to this condition the particle's Reynold's

numberRe_p, which is the ratio of inertial forces to viscous forces, and the instantaneous relative velocity V_{inst} are related as follows:

$$Re_p = \frac{\overbrace{\rho_{flow} V_{inst} d_{particle}}^{\textbf{inertial forces}}}{\underbrace{\mu}_{\textbf{viscuous forces}}} \tag{3.7}$$

What Reynold's number offer is an immediate relation between the particle's characteristics and the different flow situations. For example, high Reynold numbers indicate the existence of a turbulent flow with instabilities, like vortices or eddies, whereas low Reynold numbers will indicate smooth flow patterns. An interested reader on the association of Reynold's number and the flow situations is refereed to Hadad's and Gurka's [7] review on the influence of seeding particles parameters in PIV and PTV approaches.

3.2.2 Particle tracking and motion estimation

Following the selection of the appropriate particle characteristics and seeding density is the isolation and identification of particles. This step is crucial in the particle based methods since it can either be used as a means of computational speed and estimation accuracy increase (PIV case) or essentially to define the overall estimation accuracy (PTV case). For instance, the estimation accuracy on PTV techniques relies more on the successful tracking step since they tracking process involves only a single particle, meaning that the correct identification is crucial to ensure the validity of the result.

The most important stage in particle tracking is particle identification which is mainly based on image segmentation. To this task, the presented approaches are based on either the use of local and dynamic thresholding procedures (e.g. Stitou and Riethmuller [8]; Cavagna et.al. [9]) or particle-mask-correlation operators (e.g. Takehara and Etoh [10]) and feature selection strategies that are combined with pattern correlation metrics (e.g. Shindler et.al. [11]).

From these three method classes the one that presents the greatest interest due to its low computational cost and estimation accuracy trade-off is the last one. In this method class the most crucial step is the appropriate derivation of the features that will allow an accurate particle identification. The key feature selection condition is that the particle features must enable accurate particle identification overcoming aperture-related problems, that will lead to false motion estimates. For example, one of the most frequent

feature extraction methods is the use of corner detection based on the image intensity gradients (Harris matrix). The particle identification is based on the eigenvalues of the Harris matrix and if the minimum eigenvalue is larger than a threshold then the region is considered to contain a particle (e.g. Baker and Mathews [12]). A relevant review of feature based techniques can be found in Shi and Tomasi [13].

Motion Estimation and Validity of Estimate: The final step of a particle-based approaches involves the estimation of the motion field of the identified particles. Approaches on this task involve the application a block matching and correlation motion estimation methods with fixed (e.g. Bradley et.al. [14]; Westerweel et.al. [5]) or adjustable window sizes (e.g. Lecordier et.al. [15]; Gui and Wereley [16]), phase cross correlation (e.g. Hauet et.al. [17]; Nogueira et.al. [18]) and image deformation methods (e.g. Eckstein et.al. [19]; Eckstein and Vlachos [20]), differential (e.g. Shindler et.al [11]) as well as probabilistic methodologies (e.g. [2]).

How do we ensure the validity estimation in the particle based methods?

The validity of the estimates is determined by the displacement vector estimation success within an interrogation window. There is an immediate relation between the size of interrogation window and the particle characteristics and seeding density. According to Keane and Andrian [21] the factors that determine the estimation accuracy are: (a)tracer particle density N_I, (b)in-plane displacement amountF_I, and (c) out-plane displacement amountF_O. The latter motion cannot be appropriately determined in the image plane. Such motion in a 3D coordinate system is manifested as an object's motion across the z-axis, in the form of translation or rotation, which in the 2D image coordinate system, is reflected as a change in the object's dimensions rather than a motion. Following again Keane's and Andrian's work, a condition on the combination of the particle density and the in- and out-plane displacements within the interrogation window can be established that will allow us to control the accuracy of displacement estimation expressed as the minimization of the product $N_I F_I F_O$ (Keane and Adrian [21]; Raffel et.al. [22]), which for the case of PIV measurements in high density flows this product should than 7.

3.2.3 Particle Image Velocimetry (PIV)

As mentioned in the previous sections, we can divide the particle based methods in two method classes, (a) the Particle Image Velocimetry (PIV) method class, and (b) the Particle Tracking Velocimetry (PTV) class based on the amount of particles being tracked. Starting with the first, in PIV methods we are interesting in finding the displacement of a pattern of particles within a template window. This is done by examining a set of neighbouring locations in the next time stamp to find the most similar pattern(see Fig. 3.2).

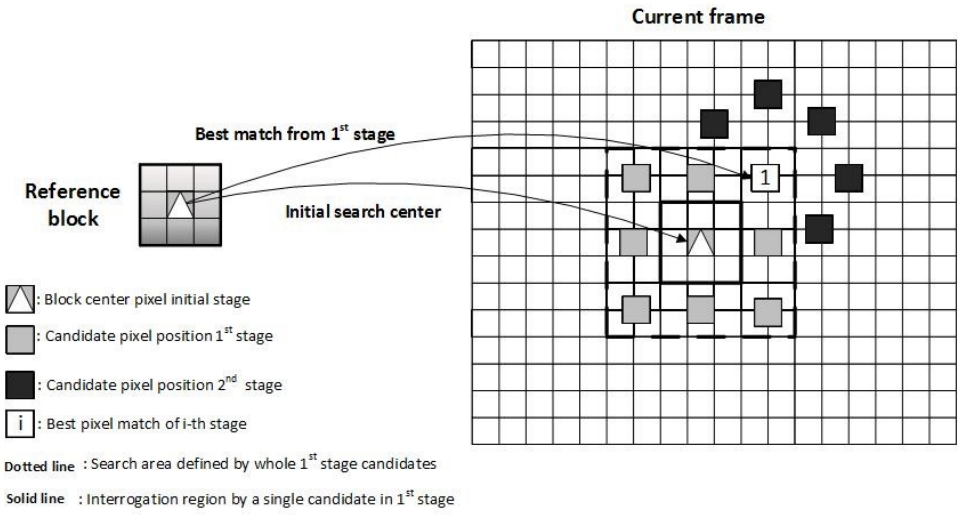


Figure 3.2: An example of a window deformation approach in a two-stage search with the interrogation window and the search area being deformable. At the 1st stage we have a square shaped window whereas in the 2nd stage becomes a circle around the best matched pixel found in the 1st stage.

The motion field is found by dividing the displacement of the image pattern along the image pair with the corresponding time interval. Existing PIV approaches can be grouped into three major categories based on the search window formulation, the space domain in which the correlation is examined and finally, the monitoring hardware used, (a) Image and Window deformation approaches, (b) Phase Cross-correlation approaches, (c) Stereoscopic camera concatenations. The last category determines hardware aspects of the particle tracking process, while it can incorporate the previous classes in the motion estimation step.

3.2.3.1 Image and Window Deformation approaches

In window based approaches, the main factors that control the estimation are the size and shape of the interrogation window. The simplest window-based motion estimation approach is to use a window with fixed size and shape and just apply a correlation coefficient as a matching criterion to extract the motion field (Bradley et.al. [14]. Such approaches, however, are constrained by the assumption that the motion is restricted within the window. This assumption makes them susceptible to out-of-plane deviations of the estimated motion vectors. A solution to this problem to adjust the window's size and shape.

Window and image deformation techniques allow the interrogation window in the second image to be deformed, in terms of rotation,size and shape. Approaches in this concept involve the use of multi-stage iterative evaluation methods that enable the adjustment of the position and the shape of the window (e.g. Lecordier et.al. [15]; Huang et.al. [23]), as shown in figure 3.2. Another approach is to deform the second image with use of a bilinear interpolation scheme or a weighting function to deform and reconstruct the second image thin order to maximize the correlation result (e.g. Jambunathan et.al. [24]; Nogueira et.al. [18]; Astarita [25]). This modification can address decision conflicts prone to large interrogation windows or truncation effects of particles in small window sizes.

3.2.3.2 Phase Cross-correlation approaches

This method class consists of approaches that apply phase filtering and correlation procedures that incorporate a series of optimized filters to the generalized cross-correlation scheme between the interrogation area and the reference pattern. The idea is that switching to the phase domain through a series of filtering procedures will enhance the standard cross-correlation result.

Such approaches however, are prune to spectral leakage due to under-sampling and window- bound discontinuity (finite window sizes) as well as aliasing effects produced by the periodic boundary constraint in the discrete Fourier case, resulting in large displacement deviations. This effect can be reduced through the incorporation of filtering procedures(prior or posterior the Fourier transform), such as spatial masking, phase filtering and Gaussian transforms of the phase correlation. The most common scheme is the Generalized Cross Correlation (GCC) that adds a series of adaptive smoothing filters

prior the Fourier transformation to reduce the effect of the background noise in the estimation of correlation. Numerous approaches have been developed on this sense, involving the incorporation of phase filters as well as Gaussian kernels prior and posterior to the FFT correlation operator (e.g. Eckstein et.al. [19]; Eckstein and Vlachos [20]; Theunissen [26]) or even the application of weighted filter series (e.g. Wernet [36]) that allows the detect of an object in the current scene based on a reference scene.

The idea is that these filters will enhance the phase term in the DFT domain that reflects the expected displacement. In the case of region matching between the reference and the candidate scene, based on a predefined displacement, the filter application will result in phase cancellation in the spatial domain and the generation of a linear phase term in the DFT domain. This will manifest as a delta−like function at the correlation peak. The position of the peak will define the amount of displacement. For example, consider figure 3.3, a filter $W(i,j)$ to the candidate region in order to detect a particle in the reference scene. In this case, the filter, known as the phase-only-filter (POF), eliminates magnitude information removing the affect of scaling, shape or size factors of the particle (Wernet [36]):

$$W(i,j) = \frac{1}{|F_2|} \tag{3.8}$$

with F_2 being the candidate region in the frequency domain.

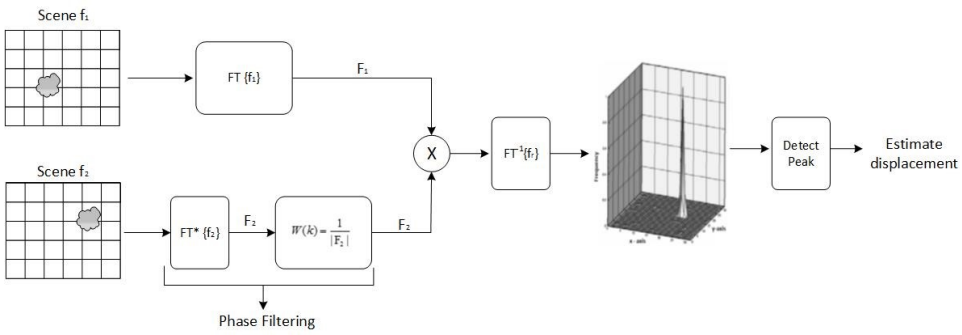


Figure 3.3: Generalized Cross-Correlation scheme using the Phase Only Filter $W(i,j)$ defined by the spectral transformation of the scene magnitude.

3.2.3.3 Stereoscopic camera concatenations

Single camera PIV systems are prune to the out-of-plane displacements. Out−of−plane motion affects the estimation accuracy due to the fact that the location of each particle does not correspond to the camera axis. Such depth perception leads to deviation between the true in-plane and the apparent motion of the particle. In order to reduce such deviations researchers have incorporated Stereo Vision in the particle estimation process. The use of a stereoscopic system can counteract the out−of−plane motion affect by simultaneously acquiring particle images from different directions. This allows a better depth perception reconstruction, since we can capture the out-of-plane motion, thus, improving the displacement estimation. The main difference of the SPIV systems from the other PIV approaches lies on the incorporation of stereo depth information in the motion estimation process. As far as the particle identification and tracking procedures used in the SPIV systems the methodologies utilized are essentially the window-based or image-deformation techniques we previously presented.

Stereoscopic PIV method can classified in two major classes based on the camera system alignment selected: (1) the translational and (2) the rotational systems.

Translational Alignment: The translational formulation, shown in figure 3.4, is produced by setting the camera pair axes parallel to each other, being symmetric to the viewed scene. Such layouts simplify the scene matching process between the two views since the deviation in the scene between the two views is manifested as a translation in the x-axis. Translational system layouts also provide well-focused images, due to the fact that each camera's optical axes is perpendicular to the illuminated area. A number of translational camera layout based SPIV approaches have been developed for liquid flow motion estimation (Prasad and Adrian [27]; Lecerf et.al. [28]; Liu et.al. [29]). For example, Prasad and Adrian [27], presented a stereoscopic approach to examine a thick liquid layer scene that uses a translational camera alignment layout in the monitoring stage, whereas in the motion estimation stage an image-shifting approach is utilized to identify and estimate the particle's in-plane and out-plane displacements.

Rotational Alignment: In the case of rotational camera alignment, as depicted in figure 3.5, the cameras are rotated over a viewing angle to the viewed scene. By rotating the camera optical center we increase field of view, thus capturing better views of

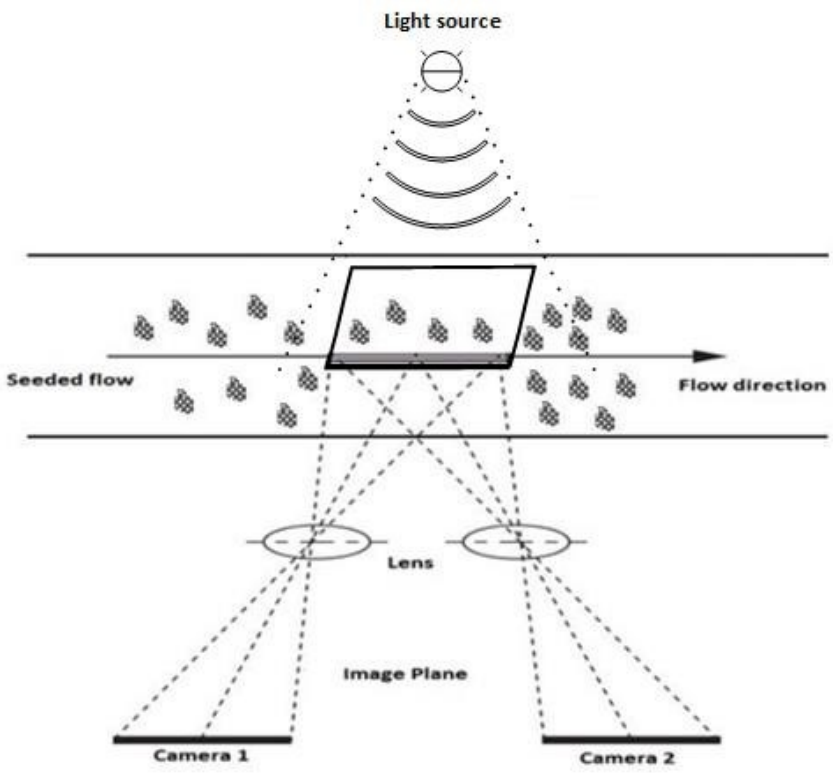


Figure 3.4: The Translational Stereoscopic camera alignment.

the out-of-plane motion, increasing the estimation accuracy due to the reduction of the perspective error. Various SPIV approaches with a rotational camera alignment have been presented(e.g. Westerweel and van Oord [30]; Hill et.al. [31]) mainly due to the accuracy increase in the out-of-plane motion that such formulations offer. However, rotational camera alignments are susceptible to errors introduced by the calibration and reconstruction procedures with the accuracy of both of them depending on the viewing angles of the cameras (Adrian and Westerweel [33]). The viewing angles of the camera greatly affect the overall estimation accuracy, since erroneous selection of them can lead to a non-uniform magnification effect for the image domain. As means of coping with this unwanted effect the image plane is further rotated under an angle bounded by the Scheimpflug condition. This condition, ensures co-linearity between the image plane, the lens plane and the particles, thus, increasing the efficiency of focus.

An interested reader on the SPIV approaches can find more details in Raffel et.al. [22] and Prasad [32].

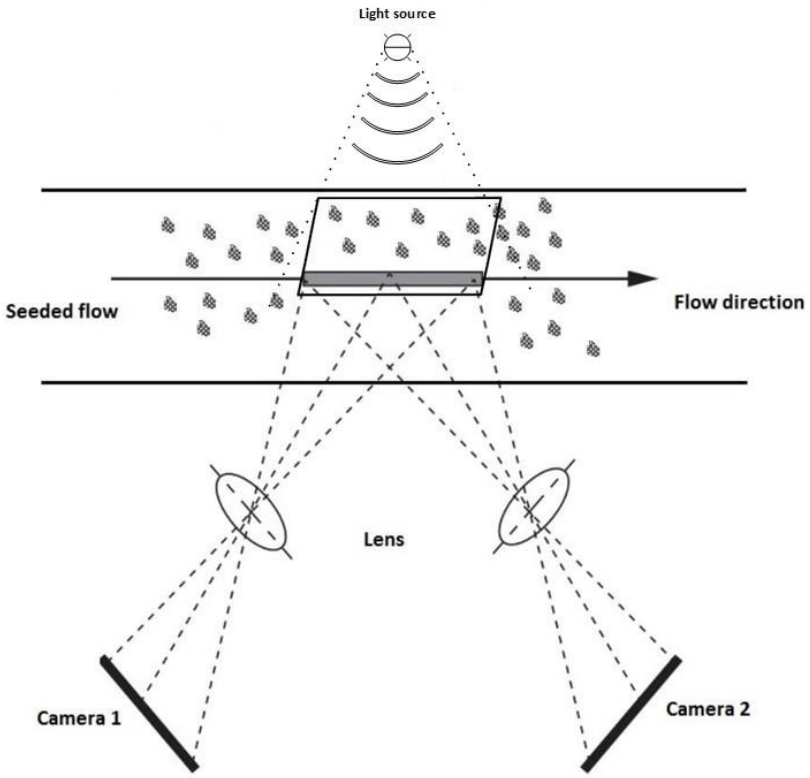


Figure 3.5: The Rotational Stereoscopic camera alignment.

3.2.4 Particle Tracking Velocimetry (PTV)

Particle Tracking Velocimetry (PTV) methods track a single particle along the frames. This means that they require low seeding densities in order to be able to identify, isolate and track the specific particle, making these crucial for the success of this method. A number of PTV approaches have been developed which can essentially be classified, as Shindler et.al. [11] suggests, into two method classes based on the number of images used in identification, isolation and tracking processes as well as the seeding density in which they can be applied: (a) multiply-exposed single images and (b) singly-exposed multiple images.

In particular, multiply-exposed single image methods are used in low seeding density sparse flows, avoiding particle occlusion and overlapping in single frames. However, singly-exposed multiple image methods are preferable in relatively higher density flows since the use of sequential frames allows better tracking of the particle. The isolation and tracking of the particle is performed by either pattern matching or feature based approaches, like Harris matrix that allow fast identification and isolation of a specific particle pattern. The tracking process can be performed by either using the initial displacement estimates extracted for the whole image domain in order to guide the particle matching or by first isolating the particle moving to a predefined region in subsequent frames in order to track it.

PTV approaches can be further grouped into two categories with respect to the motion estimation methodology used in the derivation of motion field: (a) block matching correlation-based and (b) differential-based hybrid estimation.

Correlation Methods: Correlation methods use block matching motion estimation techniques in order to derive the motion field. The matching of a particle pattern in the subsequent is achieved using correlation metrics such as the S.S.D. or M.S.E. that act as a matching criterion. For example the SSD criterion for region matching in the 2D image domain is defined as follows:

$$SSD = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left(f\left(i, j, t\right) - f\left(i + \upsilon, j + \nu, t + 1\right) \right)^2 \tag{3.9}$$

with N being the block size and $f(i, j, k)$ the pixel's intensity in the block at the k^{th} frame and (υ, ν) the optical flow field $\boldsymbol{u}$ components.

The majority of the correlation based approaches follow the first concept in the motion field and particle region association, i.e. the flow field is initially computed for the whole image domain, then identify/isolate the particle region and finally, associate and update the optical flow field only for the specified particle. A characteristic example of this method class and particle to motion field association concept is the approach presented by Stitou and Riethmuller [8], in which particle extraction is initially performed based on intensity-level thresholding as determined through the extraction of the local intensity distribution. The algorithm initially computes the velocity estimates using a block driven cross-correlation scheme followed by particle matching that associates the estimated motion field with the specified particle pattern.

Hybrid (Differential-Particle) approaches: Hybrid methods combine differential optical flow estimation approaches, with particle tracking. In this case of approaches the particle is initially identified using feature extraction techniques and a candidate region for the tracked particle in the subsequent frame(s) is defined and then differential optical flow estimation method is applied. Differential optical flow estimation methods are based on the optical flow constraint equation (equation 2.3). Such approaches combine the optical flow constraint equation with regularization and smoothness terms to define cost functions whose minimization produces the estimate of the optical flow field. For example, Shindler et.al. [11] presented an approach in which the cost function defines a relation between the candidate search regions and the reference region, in a windowed region W by combining a modified feature tracking Lucas-Kanade method with the SSD distance metric:

$$SSD(U) = \int \int_W (I_2 - I_1)^2 \, dx \tag{3.10}$$

with I_2 being the candidate image at the time stamp $t + s$ and I_1 being the reference image at the time stamp t.

This region's differential formulation, based on the optical flow constraint, in the form of partial derivatives according to each pixel intensity gradients, as Lucas - Kanade presented, goes as follows. For each pixel q_i in the window:

$$\begin{gathered} I_x(q_1)\upsilon + I_y(q_1)\nu = -I_t(q_1) \\ \vdots \\ I_x(q_w)\upsilon + I_y(q_w)\nu = -I_t(q_w) \end{gathered}$$

$$\Rightarrow \begin{bmatrix} \int\int_W I_x^2 dxdy & \int\int_W I_x I_y dxdy \\ \int\int_W I_x I_y dxdy & \int\int_W I_y^2 dxdy \end{bmatrix} \cdot \boldsymbol{u} + \begin{bmatrix} \int\int_W I_x I_t dxdy \\ \int\int_W I_y I_t dxdy \end{bmatrix} = 0 \tag{3.11}$$

This formulation is essentially a least squares problem whose solution provides the motion vector $\boldsymbol{u}$:

$$\Longleftrightarrow G \cdot \boldsymbol{u} + b = 0 \Longleftrightarrow \boldsymbol{u} = -G^{-1} \cdot b$$

with G being the Harris matrix and b the mismatch vector.

The advantage of Hybrid methods is that they combine the flexibility, speed and feature driven advantages of the differential methods to improve the particle motion field estimation. Differential approaches provide local velocity vector estimates suitable for predicting the particle's position for the successive step of particle pairing across the frames, reducing the search region range.

3.3 Probabilistic Methods

Another way to estimate the flow field of a fluid is to consider the motion vector of each fluid pixel as a random variable and thus, apply a stochastic modelling of the intensity displacement field based on a Bayesian inference scheme. The motion as a random variable is associated with a probability distribution function which enables us to use a conditional model to associate the image intensity, the unknown velocity field, the prior motion assumptions and the motion likelihood models. There have been numerous probabilistic approaches the majority whom consider Gaussian models describing the prior and likelihood information given the observations (e.g. Chang et.al. [1]; Heas et.al. [34]; Krajsek and Mester [35]).

In order to estimate the motion field $\boldsymbol{u}$ need to estimate the posterior probability of the motion field $p(\boldsymbol{u}|I)$ given a function $I(.)$ that describes image intensities. The posterior probability is determined by two factors, the likelihood or conditional probability $p(I|\boldsymbol{u})$ and the prior probability $p(\boldsymbol{u})$.

3.3.1 Conditional Probability Models

In motion estimation the conditional probability describes the observed data given the underlying motion field. The factor that defines the accuracy of such schemes is the appropriate selection of the data representation function ϕ.

One approach is to follow a stochastic formulation, in which each destination position is assigned a probability of selection. On this concept, Chang et.al. [1], used a discrete probability density function to describe the probability of displacement for a pixel in the candidate region. However, we still need to define a relation between the data and the displacement probabilities. In Chang's case the Spatio-Temporal Autoregressive (STAR)

model provides this relation:

$$I\left(x_s, y_s, t\right) = \sum_{i=1}^{D_s} A_i I\left(x_s + \Delta x_i, y_s + \Delta y_i, t + \Delta t_i\right) \tag{3.12}$$

with A_i being the displacement probability of pixel(x_s, y_s) of in the frame t to the pixel position $(x_s + \Delta x_i, y_s + \Delta y_i)$ after the time interval Δt_i.

Based on this formulation, ϕ is defined as a discrete function describing the displacements as:

$$\phi\left(\Delta x_i, \Delta y_i\right) \approx \frac{A_i}{\sum_{j=1}^{D_s} A_j} \tag{3.13}$$

In order to define a likelihood function that associates data with a motion model we combine the discrete probability function ϕ with a continuous function $g(.)$ that representing an assumption of the motion model. An example of such formulation is presented in Chang et.al. [1], where the conditional probability is defined as:

$$p\left(I|\boldsymbol{u}\right) = \prod_{x_s, y_s}^{N_s} \left(\sum_{i}^{D_s} \left(\phi\left(\Delta x_i, \Delta y_i\right) \cdot g\left(\Delta x_s - \Delta x_i, \Delta y_s - \Delta y_i\right)\right) \right) \tag{3.14}$$

In this case, the function$g(.)$ is differential function based on the pixel displacement, ϕ is the discrete function describing the data in a probabilistic formulation, D_s is the candidate displacement neighbourhood and N_s is the examined pixel's neighbourhood in the current frame.

Another approach to define the function ϕ is to incorporate the intensity conservation assumption. One representative approach of this concept was presented by Heas et.al. [34]. In their work, ϕ is defined as a linear combination of the image temporal discrepancies and an observation operator on a pixel grid, based on the optical flow constraint (equation 2.3):

$$\phi\left(\Delta x_i, \Delta y_i\right) = \frac{1}{2} \sum_{x,y \in \Omega} \left(I_t\left(x, y\right) + \nabla I\left(x, y, t + 1\right) \cdot \boldsymbol{u}\left(\upsilon, \nu\right)\right)^2 \tag{3.15}$$

with I being the intensity, $\nabla I(x, y)$ the spatial gradients of I, and I_t the temporal gradient.

In Heas et.al. [**?**], the conditional probability (likelihood) is defined in the form of a Gibbs distribution with function ϕ indicating the observed data energy:

$$p\left(I|\boldsymbol{u}\right) = \frac{1}{G_\phi} e^{-\beta_d \cdot \phi\left(\Delta x_i, \Delta y_i\right)} \tag{3.16}$$

in which G_ϕ is a normalization constant and β_d a free model parameter.

The difference between the formulation used by Chang and the of Heas lies on the fact that Heas's approach involves an energy configuration scheme in the data model instead of a probabilistic-based formulation like the one selected by Chang.

3.3.2 Prior Probability Models

As mentioned the second term of the Bayesian estimation is the prior probability formulation. This probability reflects the initial assumption on the motion model that the data follow.

In Image Processing the most preferable prior for the motion model $\boldsymbol{u}$ is a Gibbs distribution:

$$p(\boldsymbol{u}) = \frac{1}{G_{prior}} e^{-\lambda_{prior} U} \tag{3.17}$$

with prior G_{prior} being a normalization constant and U being the energy term describing the motion, usually defined as a linear combination of the partial derivatives of the optical flow field $\boldsymbol{u}$. The appropriate selection of the prior is crucial, since it acts as a smoothness factor preventing the velocity to have abrupt variations and discontinuities.

3.3.3 Optical Flow Estimation

The estimation of flow field $\boldsymbol{u}$ can be achieved through the estimation of the posterior distribution of motion $p(\boldsymbol{u}|I)$. Bayesian formulation allows the estimation of the posterior probability using the Maximum a Posteriori (MAP) rule. This rule for the case of motion field estimation is formulated as:

$$\boldsymbol{u} \propto argmax_u p(I|\boldsymbol{u}) \cdot p(\boldsymbol{u}) \tag{3.18}$$

with $p(I|\boldsymbol{u})$ being the conditional probability describing the observed data given motion field realization and $p(\boldsymbol{u})$ the prior probability describing an initial assumption for the underlying motion model.

If we now assume that the prior motion model follows an exponential distribution the MAP rule we end up with the following expression:

$$\begin{aligned} \boldsymbol{u} &= argmax_u p(I|\boldsymbol{u}) \cdot p(\boldsymbol{u}) \\ &= argmax_u p(I|\boldsymbol{u}) \cdot \alpha e^{-\lambda U} \end{aligned}$$

$$
\begin{aligned}
&\simeq argmax_u p\left(I|\boldsymbol{u}\right) \cdot e^{-\lambda U} \\
&\equiv argmax_u ln\left(p\left(I|\boldsymbol{u}\right) \cdot e^{-\lambda U}\right) \\
&\equiv argmin_u - ln\left(p\left(I|\boldsymbol{u}\right)\right) + \lambda U \\
&\equiv argmin_u L\{\phi\left(\Delta x_i, \Delta y_i\right)\} + \lambda U \\
&\equiv argmin f_{data} + \lambda \cdot f_{smooth}
\end{aligned}
$$

with L being a functional term on the conditional distribution and U being an energy based smoothness factor.

The maximization of the MAP rule for the case of exponential motion distributions is proportional to minimizing a cost function consisting of a smoothness function term and a data function term depending on the data description function ϕ. The data term f_{data} is a function describing the relation between the observed data based on the underlying motion field. Essentially, this term describes errors in the rate of change in image brightness given the estimated motion field. On the other hand, the smoothness term f_{smooth} describes the a prior assumption made about the motion model describing the flow field and imposes a smoothness constraint on the flow.

Summing up, a probabilistic formulation leads to a global motion-field estimate since the final MAP formulation relates the likelihood models of all image pixels and leads to a highly dense motion fields. However, the use of a discrete description function that assigns unique displacement probabilities to all the candidate positions in the next frame, like the one used by Chang et.al. [1] also allows the derivation of local motion estimates for pixel blocks,based only on the displacement probabilities A_i extracted for each neighborhood. A global scheme will provide smoother and more robust flow estimates. On the other, a local one will result in a loss of overall detail restricting the estimates only to an average block level, however, if we restrict to nearby object boundaries the local scheme provides more accurate results for occluded cases and boundary motion, making it ideal for particle tracking.

3.4 Differential Methods based on Fluid Properties

In differential optical flow estimation techniques, the image domain is assumed to be continuous and the optical flow is computed based on the spatio-temporal derivatives of the pixel intensity, which express the energy channeling between the frames. Global and the local schemes are also applied here based mainly on the size of implementation in the

image domain. Global approach are computed over large image regions resulting in dense optical flow fields through the minimization of a cost function based on the optical flow constraint equation applied on the image data f_{data} and smoothness and regularization constrains f_{smooth}:

$$min \int \int_{x,y} f_{data}^2 + \lambda^2 \cdot f_{smooth}^2 dxdy$$

Local methods, however, compute the optical flow field in local neighborhoods by minimizing the optical flow constraint function combined with a window based function using a least squares minimization approach.

When moving to the case of fluid flows, we can consider that the density of features in the flow is altered by the motion in a local level, thus, we can incorporate the properties of fluid mechanics, such as mass conservation, or fluid models describing fluid phenomena, like wave generation, as a constraint to justify brightness variation in the image domain:

$$\nabla \boldsymbol{I} \cdot \boldsymbol{u} + I_t = \text{Fluid Properties}$$

The addition of the fluid properties explains the divergence of the optical flow based on the undergoing affine transformation model. The optical flow field is estimated through the minimization of a cost function based on again, the optical flow constraint equation, which is now constrained by the fluid properties.

We can discretize the methodologies that follow this concept based on whether they use (a) the fluid's properties as a constraint or mathematical models or (b) mathematical models that explain the generation or the behavior of fluid phenomena, such as waves.

3.4.1 Methods based on Fluid Mechanics

Fluid properties derived from fluid mechanics, such as mass or brightness conservation assumptions, can be used as constraints in the fluid motion. One of the most popular fluid property incorporated in image-based fluid flow field extraction is the conservation of mass assumption (Wildes et.al. [37]; Nakajima et.al. [38]).

The rational behind this method class lies on the fact that the fluid's density $\varrho\left(x,y,z,t\right)$ can be associated to the 3D velocity field $V\left(x,y,z,t\right) = \left(U\left(x,y,z,t\right), N\left(x,y,z,t\right), W\left(x,y,z,t\right)\right)$ using the mass conservation assumption:

$$\nabla\left(\varrho \cdot V\right) + \frac{\partial \varrho}{\partial t} = 0 \tag{3.19}$$

The image intensity values are associated with the density the object as:

$$I(x,y,t) = \int_{z_1(x,y)}^{z_2(x,y)} \varrho(x,y,z,t)\, dz \tag{3.20}$$

with z_1 and z_2 being the object's surface boundaries. By imposing the surface boundaries on eq. (2.19), according to Fitzpatrick [53], we can associate it with the image intensity as follows:

$$\nabla_{x,y} I \cdot \boldsymbol{u} + \frac{\partial}{\partial t} I = -\left(\varrho \cdot n \cdot V_{z_1}^{z_2}\right) \tag{3.21}$$

with V being the 3D velocity estimate, ϱ the density, and $\boldsymbol{u}$ the 2D optical flow field.

The optical flow field is derived from the weighted average of the initial 3D velocity field V with the density ϱ according to the following equation:

$$\boldsymbol{u} \equiv \frac{\int_{z_1}^{z_2} \varrho \cdot V_{x,y} dz}{\int_{z_1}^{z2} \varrho dz} \tag{3.22}$$

Equation (2.21) is known as the continuity equation and indicates the relation between the 3D fluid flow and the 2D flow derived from the image under the assumption that the conservation of mass law is satisfied. The continuity equation is strengthened with additional physical and smoothness constraints, that restrict the motion field, reducing the effects of noise (e.g. [Wildes et.al. [37]; Nakajima et.al. [38]).

The flow field estimate is again found through the minimization of a cost function that has the form:

$$\boldsymbol{u} = min \int\int (k \cdot c_s + c_c) \tag{3.23}$$

with c_c being the continuity equation and c_s a smoothness constraint.

3.4.2 Methods based on the Physical Properties of Fluid Phenomena

The second approach to attribute the brightness change based on a mathematical model driven by the characteristics of specific fluid phenomena, such as wave generation or tsunami generation model. One of the most popular fluid phenomena driven models is the wave generation model (e.g. Jahne et.al. [39]; Saikano [40]).

For example, Saikano [40], presented a wave-based optical flow in which a wave generation equation is used to model the image brightness changes. The model that describes the fluid motion is a multi-directionality irregularity (MI) model of the following form:

$$I_w = \sum_{m=1}^{M} \alpha_m^* cos\left(k_m^* x \cdot cos\vartheta_m^* + k_m^* y \cdot sin\vartheta_m^* - 2\pi f_m^* t + \varepsilon_m\right) \tag{3.24}$$

with I_w being the image intensity at the pixel(x, y) as defined by the MI model,α_m^* is the amplitude, (k_m^x, k_m^y) are the wave number components, fm^* the frequency, ϑ the orientation, ε the noise and M the number of cosine functions that describe the wave.

The multi-directional irregularity (MI) model is combined with the optical flow equation to explain deviations in the estimated brightness changes based on the temporal derivative of the image intensity:

$$\nabla I \cdot \boldsymbol{u} + I_t = \frac{dI_w\left(x, y, t\right)}{dt} \tag{3.25}$$

with I being the image intensity, and I_w the expected intensity based on the MI model.

The task now is to estimate both the wave-related parameters as well as the optical flow components. To do an objective function must be defined whose minimization will lead to the desired estimation. For example, Saikano defined a combined objective function of robust logarithmic form (to accommodate for outliers and discontinuities):

$$E\left(u, v, k_m^x, k_m^y, f_m, \alpha_m, \vartheta_m\right) = \sum_{\Omega \in \mathbf{R}^2} \rho_{img_var}(e_0, \sigma) + \lambda_1 \sum_{\Omega \in \mathbf{R}^2} \rho_{wave_constraint}(e_1, \sigma) + \lambda_2 \sum_{\Omega \in \mathbf{R}^2} \rho_{smooth_constraint}(e_2, \sigma) \tag{3.26}$$

with three terms pertaining to the data from eq. (34) and two smoothness constraints, one for the wave model and the other for the optical observations. Each of these terms has the same form as: $\rho(z, \sigma) = log\left(1 + 0.5\left(z/\sigma\right)^2\right), \partial\rho/\partial e = 2e/\left(2\sigma^{2+e^2}\right), e_0 = \mid I_t + I_x u + I_y v - \dfrac{\partial H}{\partial t} \mid, e_1 = \mid u^2 + v^2 - \alpha^2\left(\gamma - f_m^2\right)^{1/2} \mid, e_2 = \mid u_x^2 + u_y^2 + v_x^2 + v_y^2 \mid, \gamma = \dfrac{3g}{16\pi h}$, g: gravity acceleration and h: water depth.

The minimization of such an objective function can be performed via optimization algorithms such as, gradient descent. The characteristic difference and advantage of specified fluid phenomena driven models is that it allows the estimation of discontinuous motion in images with inhomogeneous brightness is estimated based on a visually plausible way to reflect these expected discontinuous motion patterns. On the other hand,

methodologies based on brightness conservation can estimate rather smooth and uniform motion.

3.5 Stereo Vision

Multiple-View Geometry is an aspiring subject of Computer Vision, with rapid development over the last decades. It presents an understanding of what and how a computer system comprehends the real world. Stereo Vision is a subclass of the Multiple View Geometry subject dealing with the inference and 3D reconstruction possibilities of a viewed scene with the use of a camera pair. In order to achieve a full understanding of the 3D space through image views we need to define the number of parameters involved, the constraints between points and lines imaged in the views, and finally, how we can retrieve 3D-space points from image correspondences.

3.5.1 Projective Geometry and Estimation

Starting from the first task, i.e. the parameter derivation and estimation, we need to define the relations between the coordinate systems involved, i.e. the 2-Dimensional image plane space, the 3-Dimensional Camera space and the 3-Dimensional real world. These relations, depicted in figure 3.6, are defined as follows.

World to Camera System relation: The camera points are related with the 3-D world points based on the following transformation:

$$\mathbf{X}_c = [R|T] \cdot \mathbf{X}_{world} \tag{3.27}$$

where $\mathbf{X}_c = (x_c, y_c, z_c)$ are the camera coordinates of the point, $\mathbf{X}_{world} = (x_w, y_w, z_w)$ is the 3-D real world coordinates of the point and R, T are the rotation matrix and translation vector that relate the two coordinate systems, known as the extrinsic parameters.

World to Image plane relation: A point in the image plane is associated with its corresponding 3-D real world point through the projection matrix:

$$\mathbf{X}_{im} = K \cdot [R|T] \cdot \mathbf{X}_{world} \tag{3.28}$$

where $\mathbf{X}_{im} = (x, y)$ are the image plane coordinates of the point, and K being a 3 x 3 matrix containing the intrinsic characteristics of the system such as the camera focal

length that denote the mapping from the 3-Dimensional camera coordinate system to the 2-Dimensional image plane.

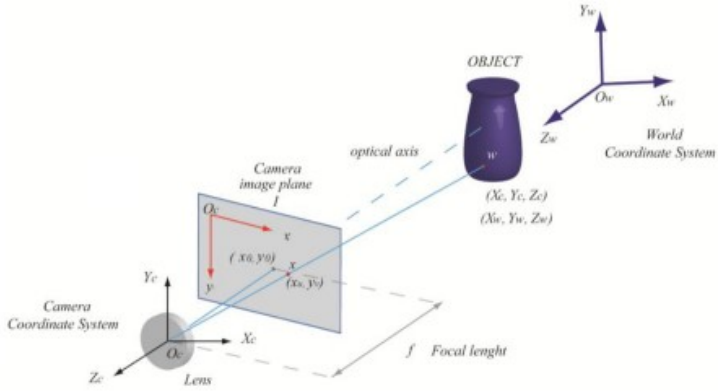


Figure 3.6: Relation between the 3-D world, the 3-D camera, and the 2-D image plane coordinate systems.

This mapping can be expressed in a matrix multiplication form with the use of homogeneous coordinates in order to express the 3D world viewed point $\mathbf{X}_{world}$ and the corresponding point $\mathbf{X}_{im}$ on the image plane:

$$\mathbf{X}_{hom.world} = (x_w, y_w, z_w, 1)^T \text{and} \mathbf{X}_{hom.im} = (x, y, 1)^T$$

Now, we can express the coordinate system association in terms of matrix multiplication as follows:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K \cdot [R|T] \cdot \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = H \cdot \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

where the image point is the 3 x 1 vector point in the homogenous image plane coordinates, K is the 3 x 3 matrix, $[R|T]$ is the 3 x 4 matrix containing the relation between the camera and real world coordinate systems, the homogeneous world point is a 4 x 1 vector point in homogeneous coordinates and finally, H is a 3 x 3 homography matrix.

(a) (b)

Figure 3.7: (a)Image with perspective distortion with the windows not being rectangular and their lines converging at a finite point, (b)Image taken from Frankfurt airport. Line parallelism is lost with lines converging also to finite point. First figure was taken from Hartley and Zisserman [42], whereas the second from wikipedia https://en.wikipedia.org/wiki/Perspective_projection_distortion.

This homography defines a geometric mapping of points from one plane to another. For the world to camera to image relations the homography matrix denotes the relation(extrinsic and intrinsic parameters) that describes the conversion of 3D real world point coordinates to image pixel point coordinates.

Plane Mapping problems and Solutions: Projection along rays through a common point defines a mapping from one plane to another. In order to say, that we have an accurate plane mapping it is evident that this mapping preserves the lines and geometrical formulations intact, i.e. a line in one plane is mapped to a line in the other (Hartley and Zisserman [42]).

However, for the case of an image we do not have this affect. Instead we end up with a projective transformation known as *perspective* projection. Such mapping results in geometrical shapes being distorted and line parallelism to be violated with parallel lines on a scene plane converging to a finite point. For example, in figure 3.7, the windows appear not to be rectangular in addition to their true nature, and line parallelism is lost as line appear to converge in the second image.

Removing projective distortion in a perspective image: In order to remove

the projective distortion we need to compute and apply to the image the inverse of the projective transformation. The way to do the computation of the projective transformation is to use point-to-point correspondences between known 3D world points and their corresponding image points. If we define a world point in inhomogeneous coordinates (x_w, y_w) and its correspondence in the image plane (x, y) then according to the projective transformation we end up with:

$$x = \frac{h_{11} \cdot x_w + h_{12} \cdot y_w + h_{13}}{h_{31} \cdot x_w + h_{32} \cdot y_w + h_{33}} \text{and } y = \frac{h_{21} \cdot x_w + h_{22} \cdot y_w + h_{23}}{h_{31} \cdot x_w + h_{32} \cdot y_w + h_{33}}$$

These equations are linear in the elements of H. In order to find the matrix H we need four point correspondences. The constraint to achieve an accurate result is that the four points must not be col-linear. By finding H we can then compute the inverse transformation of H and then apply it to the whole image to undo the effect of perspective distortion.

3.5.2 Epipolar Geometry

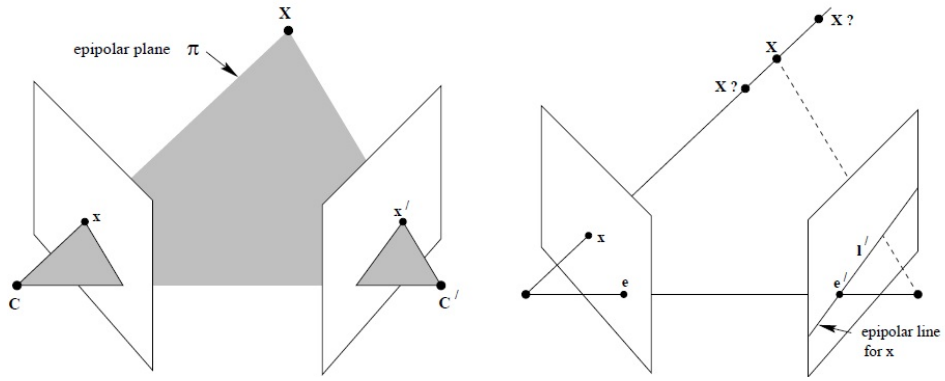
For the case of a stereo rig we need to deal with two perspective views and furthermore, we need also to define the geometry that depends only on the cameras, allowing the relation of their positions as well as their internal parameters.

The epipolar geometry is intrinsic geometry of the two camera views, depended on the cameras' internal parameters and relative pose. It essentially defines assumptions and constraints that allow the image planes of the two views to be related and guide the point correspondence search between the two views. We will now go through the basic principles of the epipolar geometry.

Epipolar plane π: The plane defined by an image point correspondence(x and x', the associate 3D space point X and the baseline of camera centres. All the parameters as shown in figure 3.8 are coplanar and the rays back-projected from the point correspondences intersect at the space point X.

How do this property facilitates the problem of searching for a correspondence?

If we know only the one of image point and search for its correspondence in the other image plane we can use the epipolar plane π to define a search area instead of searching



(a) (b)

Figure 3.8: (a)Point Correspondence, with C and C' being the camera centres of the image planes. The 3D space point X, and the image plane point correspondences images ($xandx'$) lie in the common epipolar plane π. (b) The epipolar line define a ray in which the 3D space point X must lie in order the image of the X to lie on the epipolar line l' in the the second view. Figure was taken from Hartley and Zisserman [42].

the entire image plane. We know that the ray corresponding to the unknown point lies in the epipolar plane π. If we define a line l' as the image in the second view of the ray back-projected from the known point, the unknown point (x' in the case of figure 3.8.b) will lie on the intersection of the line l' with the second image plane. This line is known as the **epipolar line**.

Epipole e**:**The point defined by the intersection of the line connecting the camera centers, known as the baseline, with the image plane.

3.5.3 Intrinsic and Extrinsic Parameters

We still have not discussed how to estimate the internal parameters of the stereo rig as well as the relative positions of the cameras between them and the viewed scene. The parameters that define these relations can be grouped into (a)intrinsic and, (b) extrinsic parameters.

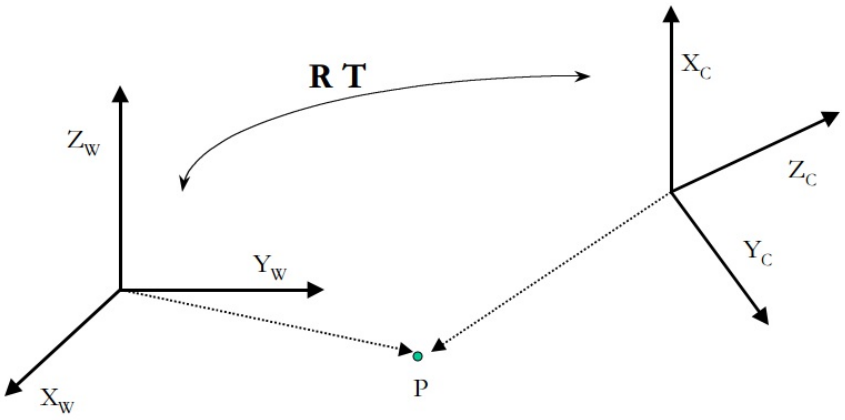


Figure 3.9: Extrinsic parameters.

3.5.3.1 Extrinsic Parameters of the Stereo rig

By the term extrinsic we refer to the transformation parameters that allow the camera and the world coordinate systems to be related, i.e. define the location and orientation of the camera with respect to the world frame.

This transformation is typically defined by (figure 3.9):

- a 3-D translation vector $T = [x, y, z]^T$ and defines relative positions of each frame.
- a 3 x 3 rotation matrix, R, that rotates corresponding axes of each frame into each other, with R being orthogonal.

As mentioned, real world, camera and image plane coordinate systems are related by projection transformations shown in equation (2.28). If we examine this relation we end up with the following pair of equations:

$$\frac{x - x_o}{f} = s\frac{x_c}{z_c} = s \cdot \frac{R_{11}x_w + R_{12}y_w + R_{13}z_w + T_x}{R_{31}x_w + R_{32}y_w + R_{33}z_w + T_z} \tag{3.29}$$

$$\frac{y - y_o}{f} = s\frac{y_c}{z_c} = s \cdot \frac{R_{21}x_w + R_{22}y_w + R_{23}z_w + T_y}{R_{31}x_w + R_{32}y_w + R_{33}z_w + T_z} \tag{3.30}$$

where (x_o, y_o) is the principal point, (x_c, y_c, z_c) is the 3D camera point's coordinates and (x, y) are the point's coordinates in the image plane, f is the focal length and s denotes the scaling ratio of the pixel spacing in the x- and y-directions for the case of unequal pixel dimensions in the CCD cameras.

Given the previous assumption and the relation presented previously we can derive the following:

$$\frac{x - x_o}{y - y_o} = s\frac{x_c}{y_c} \tag{3.31}$$

By considering only the direction of the point in the image as measured from the principle point results in an point estimate that is independent of the unknown focal length f . By combining the equations for interior (equations 2.29 and 2.30) and exterior(equation 2.27 and 2.28) orientation we obtain for the equation (2.28) the following relation:

$$(2.29)\&(2.30) \Rightarrow \frac{x - x_o}{y - y_o} = s \cdot \frac{R_{11}x_w + R_{12}y_w + R_{13}z_w + T_x}{R_{21}x_w + R_{22}y_w + R_{23}z_w + T_y}$$

By assuming that the model plane is on $Z_w = 0$ of the world coordinate system then the parameters T_z, R_{13}, R_{23}, and R_{3i} , with i=1,2,3 drop out of the equations(2.29 and 2.30) for the image coordinates:

$$\Rightarrow \frac{x - x_o}{y - y_o} = s \cdot \frac{R_{11}x_w + R_{12}y_w + T_x}{R_{21}x_w + R_{22}y_w + T_y}$$

We can observe that the estimation of the extrinsic parameters is now simplified into an 6 parameter estimation, thus simplifying the computation and data requirements. As for the overall relation between the image plane point and the 3D world point we can observe that for the case of planar targets the point X_{world} and its image plane associated point x_{im} are related by a new homography matrix:

$$x_{im} = H' \cdot X_{world} = K \cdot [R_1 \ R_2 \ T] \cdot X_{world} \tag{3.32}$$

with H' being a 3 x 3 homography matrix.

In order to estimate the extrinsic parameters we need to recover the simplified 6 parameter rotation matrix R and translation vector T. This process known as camera calibration will be presented in the following sections.

3.5.3.2 Intrinsic Parameters of the Stereo rig

As intrinsic parameters of the stereo rig we refer to the parameters affecting the relation of a camera point to its corresponding image plane projection point. These parameters are the camera characteristics, such as the focal length and the camera center point. The

matrix containing these parameters is the 3 x 3 K matrix, from equation (2.28) and in its simplest form is defined as follows:

$$K = \begin{bmatrix} f & & x_o \\ & f & y_o \\ & & 1 \end{bmatrix}$$

with f being the focal length and (x_o, y_o) the coordinates of the principal point (camera center).

In the case of a CCD camera, there is a possibility that the pixel dimensions are not the same, i.e. no squared sized pixels. In such case, the matrix K is called the camera intrinsic matrix and has the following form:

$$K_{CCD} = \begin{bmatrix} \alpha_x & & x_o^{'} \\ & \alpha_y & y_o^{'} \\ & & 1 \end{bmatrix}$$

with a_x, a_y the focal length along the x and y directions and $(x_o, y_o) = (m_x \cdot x_o, m_y \cdot y_o)$ the coordinates of the principal point (camera center).

3.5.4 Radial and Tangential Lens Distortion

The world point, image point and the optical center are not really collinear as imaged in the ideal case of the pinhole lenses. In the real non-pinhole lenses there exist a number of inevitable geometric distortions.

What is radial distortion?

Due to the spherical lens surface, a geometric distortion occurs in the radial direction. In radial distortion a point imaged at a distance from the principle point can be seen as larger/magnified (pin-cushion distortion), smaller (barrel distortion) or a mixture of both types(mustache distortion) than the perspective projection estimates (see figure 3.10).

The relation between the projected point and the ideal(non distorted) point is modeled as function of radial displacement:

$$\begin{pmatrix} x \\ y \end{pmatrix} = L(r) \begin{pmatrix} x_u \\ y_u \end{pmatrix} \tag{3.33}$$

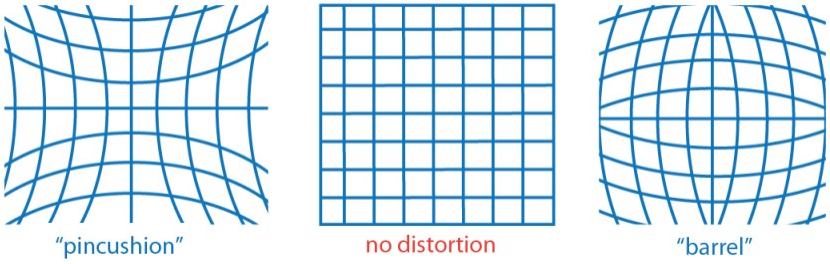


Figure 3.10: Radial distortion variations. Image source: https://en.wikipedia.org/wiki/Distortion_(optics)

where (x, y) are the distorted point coordinates, (x_u, y_u) are the ideal point coordinates and $L(r)$ is a distortion factor, defined only for positive values of r, with an approximation of it, derived using Taylor series:

$$L(r) = 1 + k_1 r + k_2 r^2 + k_3 r^3 + \cdots$$

with $(k_1, k_2, k_3, \cdots)$being the distortion coefficients and $r = \sqrt{x_u^2 + y_u^2}$. If we now consider the fact that even powers of the distance r from the principle point occur we only need to take account the first and second coefficients and ignore the others.

What is tangential distortion?

Another kind of distortion is the "thin prism" distortion, also known as tangential distortion. Tangential distortion is induced due to manufacturing imperfections of lens elements and the imperfect centering of the lens components to the camera sensor, as shown in figure 3.11.

Thin prism distortion induces both radial and tangential distortions resulting the $L(r)$ function to have the following form:

$$L(r; a) = \begin{cases} r + (\zeta_1 r^2 + \zeta_2 r^4 + \zeta_3 r^6 + \cdots)\, sin\,(\theta - a) \\ \theta + (\zeta_1 r^2 + \zeta_2 r^4 + \zeta_3 r^6 + \cdots)\, cos\,(\theta - a) \end{cases}$$

where ζ_i are the thin prism distortion coefficients, r is the radial distortion factor, θ is the observed angular component of a projected point and a is the angle between the positive

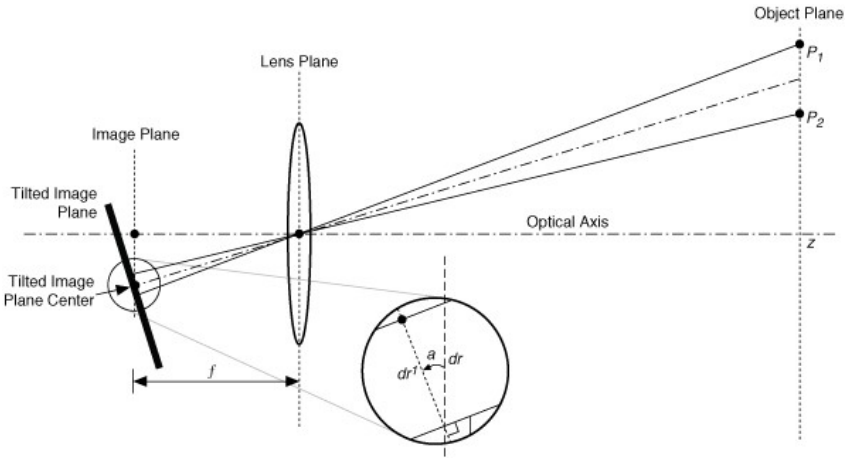


Figure 3.11: Tangential distortion, source: http://zone.ni.com/reference/en-XX/help/372916P-01/nivisionconcepts/spatial_calibration_indepth/

y-axis and the axis of maximum tangential distortion.

3.5.5 Camera Calibration

Camera calibration results in the estimation of the intrinsic and extrinsic characteristics of the stereo layout. One of the most known calibration methodology is the one presented by Zhang [41], which allows the estimation of the stereo rig relative position (extrinsic parameters) as well as the camera intrinsic parameters. Although this method is considered one of the most solid methodologies for camera calibration it only takes account the radial lens distortion and not the tangential distortion. A solution to this is to use Heikkil's and Silven's [43] intrinsic model which include two extra distortion coefficients corresponding to tangential distortion. The calibration process essentially computes the camera projection matrix P from corresponding 3-space and image entities.

We will now present Zang's method using Heikkil's and Silven's [43] intrinsic model. The basic idea behind Zang's method is that we seek correspondences between a 3D point X_{world} and its image x_{im} in order to derive the parameters of the projection transformations (extrinsic, intrinsic parameters). Given sufficiently many correspondences $X_{world,i} \leftrightarrow x_{im,i}$ the camera matrix P may be determined. In order to do so, in an unknown scene we use planar surfaces, usually a chessboard pattern, as a reference patterns

that allows easy mapping between 3D space points and the corresponding image plane ones. The restrictions for the planar surface, for the case of a chessboard is that (a) the chessboard's square size is known, (b) the chessboard sides are not even and (c) the pattern is visible in the acquired frames.

The Chessboard's role: In order to detect the chessboard pattern we deploy a corner detection procedure so as to find the corners of the chessboard. The 3D coordinates of the recognized corners are required for the estimation. In each frame(multiple frames required since we are using a non co-planar approach) the checkerboard is assumed to be coincident with the XY plane of a 3D coordinate system in which the coordinates (0,0,0) are assigned to its top-left corner. The camera centers are now reassigned by estimating the rotation and translation that minimizes their squared distances. We need to have a different aspect in the horizontal and sizes of the chessboard pattern so that the correct recognition and assignment of the coordinate layout is succeeded.

The most common approach in order to obtain the image points $x_{im,i}$, as mentioned in Hartley and Zisserman [42], is to (a)initially extract the line segments Canny edge detection, (b) to apply straight line fitting in the extracted edges and (c)find intersecting lines to derive the imaged corners.

How do we estimate the 6 parameters of the simplified planar target case and how do we compute the missing elements?

Extrinsic Parameter Estimation -Recovering the Rotation Matrix and Translation Vector: The homography matrix H' denoting the relation between the image plane point and the 3D world point is expressed as follows based on the relation equation (2.32):

$$H' = [h_1\ h_2\ h_3] = \lambda \cdot K \cdot [R_1\ R_2\ T_3]$$

with λ being an arbitrary scalar and K being the intrinsic matrix and $T : [T_x\ T_y]^T$ since T_z has been excluded in the planar simplification.

By forcing orthonormality for the first two rows of the rotation matrix R by adjusting them and re-normalizing them with a scaling factor so that they are related as:

$$R_1^{'} = R_1 + kR_2 \text{ and } R_2^{'} = R_2 + kR_1$$

and

$$R_1^{'} \cdot R_2^{'} = R_1 \cdot R_2 + k\left(R_1 \cdot R_1 + R_2 \cdot R_2\right) + k^2 R_1 \cdot R_2 = 0 \text{ and } k \approx -\left(1/2\right) R_1 \cdot R_2$$

In this case we can derive the third row of the rotation matrix R by simply taking the cross-product of the two rows. Based on the orthonormality of the two rows we can derive two following constraints for the homography matrix:

$$h_1^T \left(K^{-1}\right)^T K^{-1} h_2 = 0 \tag{3.34}$$

$$h_1^T \left(K^{-1}\right)^T K^{-1} h_1 = h_2^T \left(K^{-1}\right)^T K^{-1} h_2 \tag{3.35}$$

If we define as $A = \left(K^{-1}\right)^T K^{-1}$, a 3 x 3 matrix, and α its corresponding 6D vector, with K being the intrinsic parameter matrix as previously defined, then for each model to image homography relation we can derive the following equation:

$$h_i A^T h_j = v_{ij}^T \alpha \tag{3.36}$$

with h_i being the i-th column vector of H and

$$v_{ij} = [h_{i1}h_{j1}, h_{i1}h_{j2} + h_{i2}h_{j1}, h_{i2}h_{j2}, h_{i3}h_{j1} + h_{i1}h_{j3}, h_{i3}h_{j2} + h_{i2}h_{j3}, h_{i3}h_{j3}]^T$$

Applying the two previous relations to equation we end up with:

$$\begin{bmatrix} v_{12}^T \\ \left(v_{11} - v_{22}\right)^T \end{bmatrix} \alpha = 0 \Rightarrow V\alpha = 0 \tag{3.37}$$

With n images we end with n relations of the form of equation (2.36). The solution of this system of linear equations is the eigenvector of $V^T V$ for the smallest eigenvalue derived. By estimating V we end up estimating α and thus, finally estimating the intrinsic matrix K. Having found the estimate for K we can derive the rotation matrix and translation vector using the following relations derived from the relation (2.32) and the

orthonormallity conditions(i.e. the third row of the rotation matrix is the cross-product of the two first rows):

$$R_1 = \lambda K^{-1} h_1 \tag{3.38}$$

$$R_2 = \lambda K^{-1} h_2 \tag{3.39}$$

$$R_3 = R_1 \times R_2 \tag{3.40}$$

$$T = \lambda K^{-1} h_3 \tag{3.41}$$

with T being the translation vector (3 x 1) , $R = [R_1 \ R_2 \ R_3]$ is the (3 x 3) rotation matrix and $\lambda = 1/\|K^{-1}h_1\| = 1/\|K^{-1}h_2\|$.

Since, we are dealing with image points fused with noise(Gaussian) this approach uses a maximum likelihood inference to obtain the estimates. The estimates are obtained through the minimization of the following functional:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \|x_{ij} - \hat{x}\left(K, R_i, T_i, X_{world,j}\right)\|^2 \tag{3.42}$$

where n is the number of images with m points and $\hat{x}$ is the projection of $X_{world,j}$ to the image plane based on (2.27). The solution of the maximum likelihood problem is performed using the Levenberg- Marquardt Algorithm [44] as mentioned in the previous section.

*Intrinsic Parameter Estimation -Focal length, Principal point, Radial and Tangential distortion estimation:*Given the estimates of the Rotation matrix and the estimates for the two components of the Translation vector (T_x and T_y) we have to estimate the focal length, the T_z component, the principal point as well as the distortions and to refine the estimates based on minimization of the image error. In order to solve the previous equations we assumed that a reasonable estimate of the position of the principle point is known.

- **Focal Length and T_z**

  In order to estimate these two factors we take the cross product of equations (2.29 and 2.30), with the estimated rotations and translations (across x and y axes) R_{ij} and T_x, T_y and solve these equations for the focal length f and the translation across the z-axis T_z using one or more correspondences between the target and the image:

$$\Rightarrow \begin{cases} s\left(R_{11}x_w + R_{12}y_w + R_{13}z_w + T_x\right) f - \left(x - x_o\right) T_z = \left(R_{31}x_w + R_{32}y_w + R_{33}z_w\right)\left(x - x_o\right) & (A) \\ s\left(R_{21}x_w + R_{22}y_w + R_{23}z_w + T_y\right) f - \left(y - y_o\right) T_z = \left(R_{31}x_w + R_{32}y_w + R_{33}z_w\right)\left(y - y_o\right) & (B) \end{cases}$$

- **Radial and Tangential Distortion**

  Based on the intrinsic model defined by Heikkil's and Silven's [43], the distortion model between the true (distorted) image coordinate and the undistorted image coordinate (produced by the pinhole geometry) is defined as a function of both radial and tangential distortions:

$$\begin{cases} x_u + D_x + d_x = x \\ y_u + D_y + d_y = y \end{cases} \tag{3.43}$$

  where (x, y) is the distorted image coordinate on the image plane ,and (x_u, y_u) is the undistorted image coordinate and D_x, D_y is the induced radial distortion and d_x, d_y is the induced tangential distortion with:

$$\begin{aligned} D_x &= x_u\left(k_1 r^2 + k_2 r^4 + \cdots\right), \\ D_y &= y_u\left(k_1 r^2 + k_2 r^4 + \cdots\right), \\ r &= \sqrt{x_u^2 + y_u^2}, \\ d_x &= 2k_3 x_u y_u + k_4\left(r^2 + 2x_u^2\right), \\ d_y &= k_3\left(r^2 + 2y_u^2\right) + 2k_4 x_u y_u \end{aligned}$$

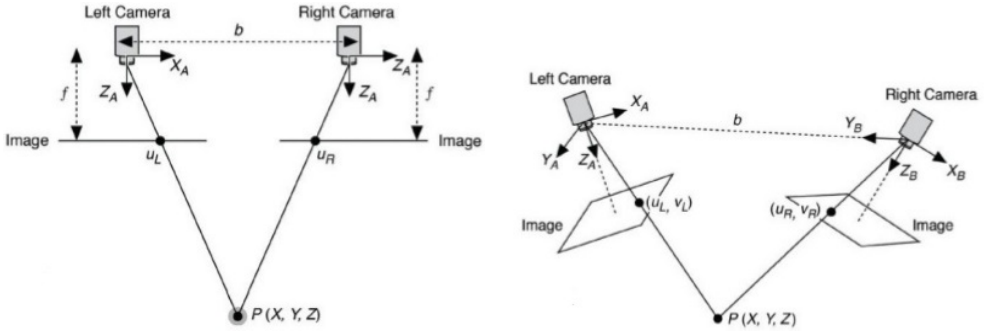
where k is the 5 x 1 vector containing the distortion coefficients(radial and tangential) The estimation of the intrinsic parameters is incorporated to the Maximum Likelihood inference model allowing the estimation of the complete set of parameters by minimizing the following functional:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \left\| x_{ij} - \hat{x}\left(K, k_1, k_2, k_3, k_4, R_i, T_i, X_{world,j}\right) \right\|$$

The minimization is again performed with the Levenberg- Marquardt Algorithm as mentioned previously.

3.5.6 Stereo Alignments and Image Plane Relation

The continuous debate in a stereo rig formulation is whether a parallel or a convergent layout will be selected. Each one has its advantages and disadvantages. In a parallel rig the cameras are placed parallel to each other separated by the inter-axial distance and both are aimed straight ahead, as illustrated in figure 3.12).(a). A parallel rig requires simpler transformations in order to move from the physical to image plane coordinate systems as well as to relate the two image planes together. In a parallel rig the the viewed scene in each camera differs in a form of translation in the x-axis only. The biggest problem in a stereo rig is to find a way to relate the views of the two cameras, thus a simple translation in one axis allows us to easily find point correspondences between the image views. A parallel layout provides valid points more accurately, thus, allowing a more accurate 3D scene reconstruction with denser depth information fields compared to a convergent layout, and is also unaffected by keystone distortion ** that is more common in a convergent stereo rig.
However, parallel layouts provide less information about the depth perception, since the field of view is constrained. This fact is due to the point of convergence of the two views, which in a parallel stereo rig happens to be the infinity. When the convergence point is infinity, then depth information of the scene is lost and there is no focus. Also, this introduces a number of unmatched regions in the views especially in the edges of the scene.The solution to this problem is to shift the two views horizontally to place the point of convergence wherever desired. This is an easy task since the views only differ in the x-axis.
On the other hand, convergent layouts, use wider viewing angles and depth perception, since the point of convergence is not the infinity as in the parallel case but in front of the viewed scene. This means that everything in front is pushed in front of the view anything behind is pushed back into screen space, whereas, an infinity point of convergence does not produce this effect and, as mentioned, requires post processing. Nevertheless, convergent layouts introduce keystone distortion, when the viewing angles and viewed scene's dimensions are not selected appropriately, which can lead to inaccuracies during the 3D scene reconstructions. Moreover, since the cameras are not aligned further pre-processing



(a) (b)

Figure 3.12: (a)Parallel stereo rig, (b) Convergent stereo rig. Figure was taken from http://www.slideshare.net/RobinColclough1/viva3d-stereo-vision-user-manual-en-201606.

must be made in order to match the views (x and y axis misalignments). This is known as image rectification. As mentioned this process is not required in parallel rigs due to the fact that the views only differ in x-axis information.

***Keystone Distortion*: an effect produced when two images are not parallel, due to the projection of an image onto a surface at an angle which results in a distortion on the image dimensions, e.g. make a square look like a trapezoid.

3.5.6.1 Image Rectification

In the case of a convergent stereo rig the images taken by the camera pair are mismatch in both x and y axes. In order to simplify our search for point correspondences during the relation of the viewed scene to its corresponding image plane view we need to apply geometrical transformations that changes a general camera configuration model with non-parallel epipolar lines (convergent layout) to the canonical one (parallel image planes) that will allow the mismatches to be present only in the x-axis(see figure 3.13).

The rectification procedure is based on the characteristics and constraints of the stereo rig and the common plane of projection π defined by the two camera centers and their corresponding image planes. The characteristics of the epipolar plane π defined after the rectification process will be:

- Epipolar lines(l and l' in the figure 3.13) become collinear and parallel
- Exclude y-axis from the disparity estimation step

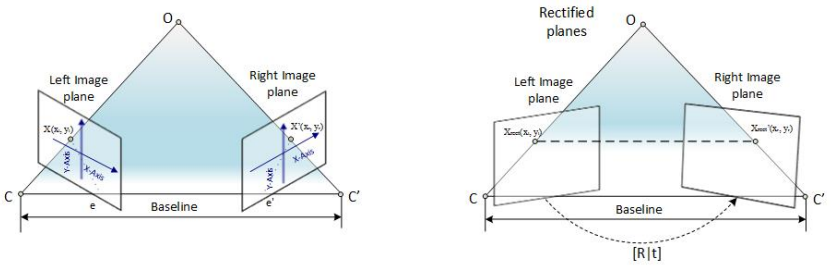


Figure 3.13: The rectification process of transforming the image pair, with C, C' being the principal points, and [R|t] the transformations relating the two planes.

- Any disparities will be parallel to x-axis
- Epipoles $\rightarrow \infty$

The rectification procedure, as previously mentioned, is the process of computing the image transformations resulting the conjugated epipolar lines (denoting the corresponding points) to become collinear and parallel to horizontal image axis. In order to define the require transformations that allow the relation of the two images planes and the world coordinate system let's remember the relation between the 3D world point coordinate vector $X_{world} = (x_w, y_w, z_w)^T$ and the corresponding coordinate vector in the camera reference coordinate point $X_{camera} = (x_c, y_c, z_c)^T$ is defined according to equation(2.27).

In the case of the camera pairs, the aim is to relate the two image planes in order to associate corresponding points but also to define the relations the image planes and the physical world. In this case the camera projection transformation for each camera are the following:

$$X_{left} = P_{left} \cdot X_{world} \text{ and } X_{right} = P_{right} \cdot X_{world} \tag{3.44}$$

with P_i being the projection matrix of each camera relating each image point X_i of the each camera to the world point X_{world}.

What we need to do now is to define the appropriate rotational and translational transformations that will form the projection matrix of the left image plane in accordance with the one of the right camera's, i.e. rectify the planes:

$$P_{left} = R \cdot P_{right} + T \tag{3.45}$$

with R, T being the rotational and translational transformations.

One of the most known image plane rectification methods is the one introduced by Trucco and Verri [45]. The method is based on the fact that the extrinsic as well as the intrinsic parameters of the stereo rig are known and that the stereo rig is bound/constrained based on the following assumptions: 1)the origin of the reference frame is the principal point and 2) the focal length is known. The methodology goes as follows:

1. Estimate the intrinsic and extrinsic parameters of both images. For example, we can use Zang's [41] approach (known and well tested approach) to recover the interior orientation, the exterior orientation, the power series coefficients for distortion, and an image scale factor. This is done by using linear least-squares fitting methods.

2. Calculate the relation transformations, i.e. the rotation and translation require in order to match the the planes of the two cameras:

$$R = R_{right} \cdot R_{left}^{T} \text{ and } T = T_{left} - R^{T} \cdot T_{right} \tag{3.46}$$

   where R_i, T_i are the rotation matrix and translation vector of the i-th camera, $i = [left, right]$.

3. Construct three mutually orthogonal unit vectors e_1, e_2 and e_3. The first vector e_1 is given by the epipole. Since the origin coincides with the image center the vector e_1 has the same direction with the translation and can be defined as:

$$e_1 = \frac{T}{\|T\|}$$

   For the second vector e_2 we have the orthogonality constraint. By taking the cross product of e_1 with direction vector of the optical axis and normalizing it, we derive the vector e_2:

$$e_2 = \frac{1}{\sqrt{T_x^2 + T_y^2}} \left[-T_y \; T_x \; 0\right]^T$$

   Now, again using the orthogonality principle between the three vectors we can derive the vector e_3 as the cross product of the vectors e_1 and e_2:

$$e_3 = e_1 \times e_2$$

Build the orthogonal matrix R_{Rect} that will be used to define the new rotation matrices and translation vectors of the two camera's that will define the new Projection matrices:

$$R_{Rect} = \begin{bmatrix} e_1^T \\ e_2^T \\ e_3^T \end{bmatrix} \tag{3.47}$$

4. Set $R_{left} = R$ and $R_{right} = R \cdot R_{Rect}$. The projection matrices of the image pair can are associated based on the R and T computed in a relation similar to the one relating the world coordinate system and the camera coordinate system:

$$P_{left} = R^T P_{right} T \Rightarrow R_{Rect} P_{left} = R_{Rect} R^T P_{right} + R_{Rect} T \text{ and } R_{Rect} T = \begin{pmatrix} \|T\| \\ 0 \\ 0 \end{pmatrix}$$

5. For every point of the left image calculate the rectified point as follows:

$$x_{im,left} = [x \ y \ f]^T \Rightarrow R_{Rect} x_{im,left} = [x^{'} \ y^{'} \ z^{'}]^T \Rightarrow x_{im,leftRect} = \frac{f}{z^{'}} [x^{'} \ y^{'} \ z^{'}]^T$$

and do the same for the right camera points. But these points are in camera coordinate system (not in pixels). To associate them with pixel coordinates:

$$x - o_x = \frac{f_x}{f} - x_c \text{ and } y - o_y = \frac{f_y}{f} - y_c$$

with (x_c, y_c) being the camera center coordinates, f the focal length and (o_x, o_y) the coordinates of the center of projection.

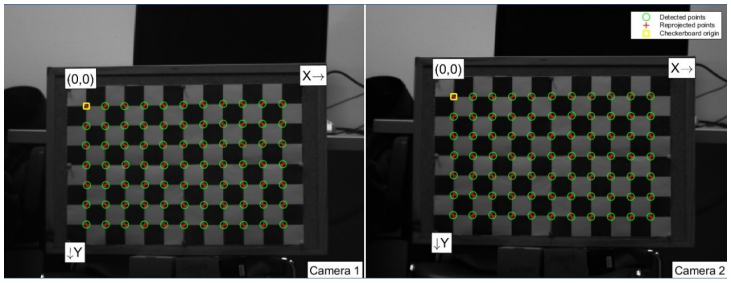
Perform the last step backwards, which means for each pixel in the rectified image, we need to find the correspondent point in the original image so that we don't end up with holes in the rectified image.

The final relations between the rectified image points and the original image points are the following:

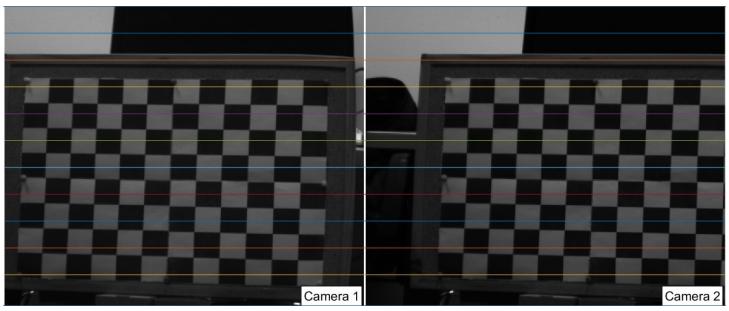
$$x = \left(x^{'} - o_x\right) \frac{z^{'}}{f} R_{11} + \left(y^{'} - o_y\right) \frac{z^{'}}{f} \frac{f_x}{f_y} R_{21} - R_{31} \frac{z^{'}}{f} f_x + o_x \tag{3.48}$$

$$y = \left(x^{'} - o_x\right) \frac{z^{'}}{f} \frac{f_x}{f_y} R_{12} + \left(y^{'} - o_y\right) \frac{z^{'}}{f} R_{22} - R_{32} \frac{z^{'}}{f} f_y + o_y \tag{3.49}$$

An example of a pre-rectified and the rectified image pair for the chessboard pattern is shown in figure 3.14.



(a)



(b)

Figure 3.14: (a) Chessboard view from a convergent stereo rig prior rectification with disparities present in both x and y axis, (b) Rectified Chessboard view with disparities only present in the x-axis.

3.5.7 3D Structure Reconstruction via Triangulation

Having related the two views and the world system we can determine a point in the 3D space given its projections onto two, or more, images, a process known as triangulation. In order to solve this problem it is necessary to know the parameters of the camera projection function from 3D to 2D for the cameras involved, in the simplest case represented by the camera matrices. Having estimated the 3D coordinates of each point of the scene we can reconstruct the scene.

How do we find the 3D space point?
In order for point correspondences be associated with a 3D space point the must satisfy the epipolar constraint. However, an image pair point correspondence observation consists of noisy point correspondences which does not in general satisfy the epipolar constraint.

What is the epipolar constraint? The epipolar constraint $x^{'}Fx = 0$ denotes that the point $x^{'}$ in the second image lies in the $F \cdot x$ and the two rays back-projected from image points x and $x^{'}$ lie in a common epipolar plane passing through the two camera centers, with F being the fundamental matrix.

The idea behind this constraint is derived from the fact that if the projection point x is known, then the epipolar line $e \cdot x^{'}$ is known and the point $X_{wor;d}$ projects into the right image, on a point $x^{'}$ which must lie on this particular epipolar line, as shown in figure 3.8.(b). This means that for each point observed in one image the same point must be observed in the other image on a known epipolar line. This provides an epipolar constraint which corresponding image points must satisfy and it means that it is possible to test if two points really correspond to the same 3D point.

Fundamental Matrix, role and its use in the epipolar constraint:

The fundamental matrix is the algebraic representation of epipolar geometry. It is the projective mapping from points to lines describing the epipolar constraint we previously mentioned. More specifically, the set of points x and $x^{'}$ are both images of the 3D point X_{world} in the plane, i.e. they are projectively equivalent, since they are each projectively equivalent to the planar point set X_{world}. Thus, there exists a 2D homography mapping the one with other. Now, given the point $x^{'}$ in the second image, the epipole $e^{'}$ and the epipolar line $l^{'}$ passing through $x^{'}$ for which we have $l^{'} = e^{'} \times x^{'}$. Then, since the points x and $x^{'}$ are associated through a homography matrix H we can express the epipolar line $l^{'}$ as:

$$l^{'} = e^{'} \times H \cdot x = F \cdot x$$

with $F = e^{'} \times H$, being the fundamental matrix.

- **Camera projection matrices and fundamental matrix association**

  Given the camera pair projection matrices P and $P^{'}$ of the left and right camera respectively. The association of the 3D world point X_{world} with the image point x and its projection matrix is $P \cdot X_{world} = x$. The ray back-projected from the world point X_{world} to the image plane into x is the line formed by the two known points, the camera center C (for which $P \cdot C = 0$) and the point $P^{+}x = P^{T}\left(PP^{T}\right)^{-1} \cdot x$, which belongs to the ray since it projects to x, $P(P^{+}x) = x$. The line joining the

two points is:

$$X_{world}(\lambda) = P^{+}x + \lambda C \tag{3.50}$$

with P^{+} being the pseudoinverse of P, $PP^{+} = X$ffl, C is the camera center and X)) is a normalization scalar.

For the second camera, with a projection matrix $P^{'}$ these two mentioned previously are imaged at $P^{'}C$ and $P^{'}(P^{+}x)$. The epipolar line $l^{'}$, joining these two points, is:

$$l^{'} = \left(P^{'}C\right) \times \left(P^{'}P^{+}x\right)$$

The point $P^{'}C$ is the epipole $e^{'}$ in the second image and thus, the previous equation is written as:

$$l^{'} = e^{'} \times \left(P^{'}P^{+}\right)x = Fx$$

meaning that the fundamental matrix, given that the projection matrices are known is equal to $F = e^{'} \times \left(P^{'}P^{+}\right)$.

- **Fundamental Matrix Properties(as presented in Zisserman [42])** (1) The epipole: "For any point x (other than e) the epipolar line $l^{'} = F \cdot x$ contains the epipole $e^{'}$. Thus $e^{'}$ satisfies $e^{'T}(F \cdot x) = \left(e^{'T}F\right)x = 0$ for all x. It follows that $e^{'T}F = 0$, i.e. $e^{'}$ is the left null-vector of F. Similarly $F \cdot e = 0$, i.e. e is the right null-vector of F". (2) For epipolar lines: $l = F\cdot^{'}$ and $l^{'} = F^{T} \cdot x$.

3.5.7.1 Triangulation Methodology

Going back to the derivation of the 3D space points we must find the best point correspondence estimates that minimize a geometric error subject to the epipolar constraint given a fundamental matrix F in order to find the correct correspondence. Based on Hartley and Zisserman [42], the optimal triangulation method first corrects the point correspondences found initially, in order to find the best corresponding points satisfying the epipolar constraint, and then compute the 3D world point correspondence based on the DLT method.

- **First Step: Point correspondence correction**

  The objective of this step is given a measured point correspondences $x \leftrightarrow x^{'}$, and a fundamental matrix F, compute the corrected correspondences $\hat{x} \leftrightarrow \hat{x}^{'}$that minimize the geometric error(below) subject to the epipolar constraint:

$$C\left(x, x^{'}\right) = d\left(x, \hat{x}\right) + d\left(x^{'}, \hat{x}^{'}\right), \ \hat{x}F\hat{x}^{'} = 0 \tag{3.51}$$

  where $d()$ is the Euclidean distance. The point correspondence correction is based on the assumption that only one of the corresponding points lies at an epipole leading the 3D space point to coincide with the other camera center. The overall algorithm used to achieve the identification and correction of the point correspondences is summarized in Hartley and Zisserman [42], chapter 12, algorithm (12.1) - *until step x.*

- **Second Step: 3D world space point estimation** After the correction of the point correspondences we can use the estimates to find the 3D world point X_{world} estimate using the Homogeneous method (DLT). The method uses the corrected point correspondences, i.e. $\hat{x}$ and $\hat{x}^{'}$.

  Homogeneous method(DLT): We know that each pair of image points of the image pair used is associated with its 3D space point based on the equations $x_{im} = P \cdot X_{world}$ and $x_{im}^{'} = P^{'} \cdot X_{world}$. These equations can be combined into a form $AX = 0$, which is an equation linear in X, with A a matrix being of the following form:

$$A = \begin{bmatrix} x \cdot p^{3T} - p^{1T} \\ y \cdot p^{3T} - p^{2T} \\ x^{'} \cdot p^{'3T} - p^{'1T} \\ y^{'} \cdot p^{'3T} - p^{'2T} \end{bmatrix}$$

  with $x_{im} = (x, y), x_{im}^{'} = \left(x^{'}, y^{'}\right)$ the point correspondences in each image, and $P = \left[p^{1T}\ p^{2T}\ p^{3T}\right], P^{'} = \left[p^{'1T}\ p^{'2T}\ p^{'3T}\right]$ the projection matrices.

  So, all we need is to use the matrix A to estimate the 3D space point X_{world}. The algorithm goes as follows (as presented in the Hartley and Zisserman's book, chapter 11).

The overall method is summarized as: The task is given $n \geq 4$ point correspondences, and the matrix A, determine the vector X so that $AX = 0$, with X being the vector describing the 3D space point.
For each correspondence compute the matrix A_i, as previously shown.

1. Assemble the $n(2 \times 9)$ matrices A_i into a single $2n \times 9$ matrix A.
2. Obtain the SVD of A, the SVD is a factorization of A as UDV^T, where U and V are orthogonal matrices, and D is a diagonal matrix with non-negative entries. The unit singular vector corresponding to the smallest singular value is the solution X_i. Specifically, if $A = UDV^T$ with D diagonal with positive diagonal entries arranged in descending order down the diagonal, then X is the last column of V.
3. The matrix X is determined.

Thus, we have found the 3D world space points that correspond to the points of the two images planes.

3.5.7.2 Reconstruction Ambiguities

We will present the reconstruction capabilities of a scene based on the knowledge available about the scene's placement with to a 3D coordinate frame, as well as the parameters of the monitoring system(see figure 3.15).

- **Projective transformation ambiguity:** if we don't know anything about the intrinsic and extrinsic parameters of the monitoring system, then, we can we can express the ambiguity of reconstruction with an arbitrary projective transformation. A projective transformation will preserve intersection and tangency but will not preserve angles, ratios of length or volume of an object.

- **Affine transformation ambiguity:** if we now know the focal length of the cameras and the cameras are associated through a simple translation, then the reconstruction can be expressed with an affine transformation. An affine transformation will preserve parallelism and volume ratios, however, it will not preserve angles and ratios of length.

- **Similarity transformation ambiguity:** If we have a fully calibrated camera pair, then reconstruction can reach up to a similarity transformation. A similarity

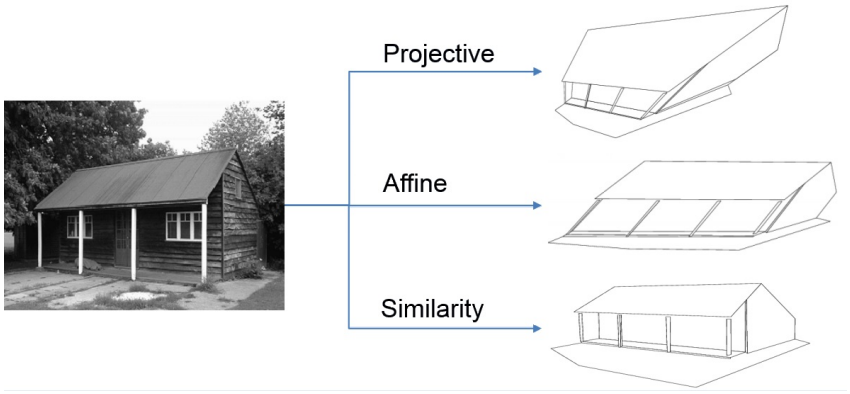


Figure 3.15: The reconstruction transformations for the case of the building view. Image transformed based on an image presented in Hartley and Zisserman [42].

transformation will preserve parallelism, angles, ratios of volume and length. However, we still deal with scaled ratios of length, in order to derive actual information about the real world dimensions of the object we will need additional information about the real dimensions of the scene, thus, leading to an Euclidean reconstruction (*metric reconstruction*!

Metric transformation: This transformation preserve the full extent of information that we can extract from a scene, parallelism, angles, length and volume. In order to move from another reconstruction scale to the metric scale additional information about the scene are needed. This information can be provided in the form of a constraint on line parallelism that can correct the plane of estimation or ground truth points (simplest and most effective way). For example, in the case where ground truth points are available then the relation between a ground truth point X_{grP} and the estimated 3D point X_{est}, is through a 3 by 3 Homography matrix H (Hartley and Zisserman [42]):

$$X_{grP} = H \cdot X_{est}$$

which can be moved to an image point relation reforming the image point and world correspondence equation(2.28) as:

$$X_{im} = P \cdot H^{-1} \cdot X_{grP} \tag{3.52}$$

An example of the impact of the use of ground truth points in the case of a projective reconstruction is shown in figure 3.16.

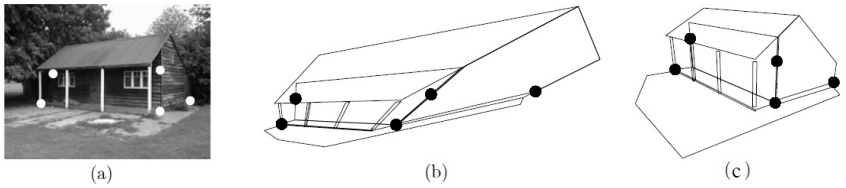


Figure 3.16: The The projective reconstruction moved to metric through the use of five (or more) world points: (a) the five points placed on the initial scene and, (b) the reconstruction after the point mapping. Image taken from Hartley and Zisserman [42].

Chapter 4

Our Approach

Our approach is a new formulation of an image-based river monitoring system that is unbound by the requirement of particles for the generation of a velocity flow estimate as well as, of the task of control point selection for the estimation of the real surface velocity of the river. Our method utilizes a stereo camera layout to derive the necessary relation between the physical and image coordinate systems. The estimation of the optical velocity of the fluid is performed using a stereo probabilistic framework for the computation of the optical flow field. To increase the accuracy of the estimation as well as to remove erroneous or unwanted motion vectors, image segmentation and machine learning classification methods are incorporated. Our approach can be summarized in the following steps:

1. **Stereo Layout**:Use the stereo layout to derive the relation between the 3D physical and 2D image coordinate systems and formulate a region of examination with known real world dimensions.

2. **Optical Flow Estimation**: Combine a probabilistic optical flow estimation methodology with the additional information provided by the stereo (layout) to estimate the optical flow field of the fluid employing the entire image domain of the stereo image pair.

3. **Coordinate System Relation for Velocity Estimation**: Associate the estimated motion field with the corresponding 3D physical velocities based on the 3D physical coordinate change of the 2D image points as defined by the 2D motion vector. In this way the perspective distortion introduced by the transformation

from the 3D world to 2D image domain is removed.

4. **Velocity Estimate Validation**: Consider the average 3D velocity estimate as constant over the monitored area and validate the estimation by examining whether an existing particle in the flow, assuming to have the same velocity, travels along the examination area at the expected number of frames. By validating our estimate we validate the fact that the estimated velocity accurately expresses the fluid's motion.

4.1 Stereo Monitoring Layout

As mentioned, existing monitoring systems make use of a single camera set up. The relation between the physical and image coordinate systems is established by solving an eight parameter transformation system, which given the appropriate amount of data (at least 8 control points) leads to a system of linear equations. However, such formulations require the prior appropriate selection of known control points which in accordance with the horizontal viewing position assumption leads to harsher geometrical reconstructions of the scene (affine reconstruction).

To overcome these disadvantages and reach to more detailed scene reconstructions, we propose the use of a stereo layout that can provide world and image plane mapping without the use of predefined control points, as well as providing information about the depth map of the scene allowing even crude 3-Dimensional scene reconstructions. In our monitoring formulation we have selected a calibrated monitoring approach since it allows the computation of both the stereo rigs' characteristics and the distortion values, minimizing projection ambiguities and thus, strengthening the coordinate system relation. The use of a stereo rig apart from removing the need of control points in order to relate the physical and image coordinate systems, provides additional scene information, such as depth perception and multiple scene views, that can be utilized in the processing part to increase the accuracy of estimation e.g. stereo motion estimation for the flow or increase field of view by view stitching (*panorama*).

As mentioned in Chapter 3, the continuous debate in a stereo rig formulation is whether a parallel or a convergent layout will be selected. Each of these has its advantages and disadvantages. A parallel layout requires simpler transformations in order to move from the physical to image plane coordinate systems and provides more valid points for 3D scene reconstruction i.e. denser depth information field compared to a convergent

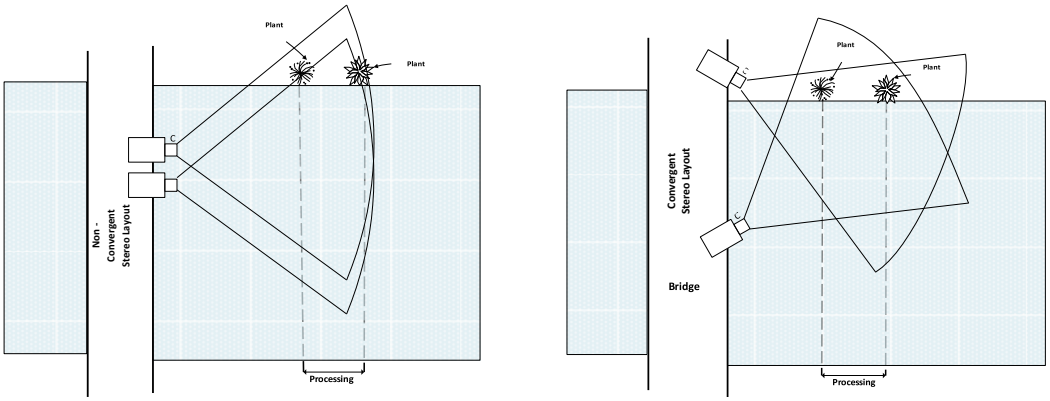


Figure 4.1: (a)Non Convergent Stereo Rig, (b) Convergent Stereo Rig, placed on a bridge.

layout. On the other hand, parallel layouts provide less information about the depth perception. Convergent layouts, have better viewing angles and depth perception but, introduce keystone distortion which reduces the number of valid points in both image planes, and thus, leading to sparser scene reconstructions. The answer to this problem in the river monitoring case is that the selection is river dependent. Streams or rivers with small width will allow a parallel layout place in the center of the bridge as well as, a convergent layout (see figure 4.1). However, for the case of a river with large width only a convergent layout will be feasible due to the fact that the parallel layout has a restricted distance between the camera pair ($\approx 5.5cm$) resulting in a restricted field of view, and thus, in our system formulation a convergent layout was selected.

However, as mentioned in Chapter 3, convergent stereo layouts require the association of camera pair positions, which in our case differ in x and y axes(in non-convergent they differ only in x-axis), as well as the stereo rig with the physical 3-Dimensional world system. In the following subsections we will define these relations, based on the theoretical background presented in Chapter 3.

4.1.1 Physical to Image plane coordinates

The role of the stereo rig is to connect the physical and image plane coordinate systems so that an association between the 3-Dimensional viewed scene and the 2-Dimensional image can be defined, providing us with a depth map estimate of the scene referenced to the stereo layout position. This coordinate system relation allows us to define a region of known dimensions that will provide the information required to perform the transformation of the image-based velocity estimate to its corresponding real world velocity approximation. The relation between a point in the physical $X_world = (x_w, y_w, z_w)$ and a point in the image plane $x_{im} = (x, y)$ coordinate systems is given by:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K \cdot [R|T] \cdot \begin{bmatrix} x_{world} \\ y_{world} \\ z_{world} \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} f & & x_o \\ & f & y_o \\ & & s \end{bmatrix} \cdot \begin{bmatrix} R_{11} & R_{12} & R_{13} & |T_x \\ R_{21} & R_{22} & R_{23} & |T_y \\ R_{31} & R_{32} & R_{33} & |T_z \end{bmatrix} \cdot \begin{bmatrix} x_{world} \\ y_{world} \\ z_{world} \\ 1 \end{bmatrix}$$

$$\Rightarrow x_{im} = P \cdot X_{world}$$

with K being a matrix containing the intrinsic characteristics of the camera, required for the relation between the camera and the image plane coordinate systems, $[R|T]$ a matrix denoting the extrinsic parameters of the system, i.e. the rotation (3 x 3 matrix) and translation (3 x 1 vector) required to match the world and camera coordinate systems and P being the projection matrix, defining the geometric mapping of points from one plane to another.

This equation expressed the correspondence relation between the image points and the world points. In order to estimate the intrinsic parameters of the system (if not known), such as the focal length f or the principal point coordinates $(x_o, y_o$, as well as the extrinsic characteristics of the camera to world relation, i.e. the relative position of the stereo layout in accordance to the viewed scene, we follow a calibration process(as presented in Chapter 3).

The estimation of both intrinsic and extrinsic parameters of the system allows us to

reduce the projection ambiguity between the two coordinate systems (up to similarity reconstruction or Euclidean, given additional scene information), thus acquiring more detailed point mapping for the final point displacement association.

4.1.2 Image Plane relation in the camera pair

Having related the physical and image coordinate systems for each camera the next step is to relate the camera pair together. This process involves the disparity map estimation. For the case of a non-convergent layout the image planes are parallel on the y-axis, meaning that the difference between a point in the first camera is just a translation in the x-axis, i.e. a point at the position (x, y) in the first image will be located in the position $(x + d_i, y)$ in the second image. For the case of a convergent layout in order to relate the image planes we need first to rectify them (as has been shown in figure 3.13 so that the points can be associated with a translation in the x-axis only, and finally compute the disparity map.

The relation between the 3D coordinate point $X_{world} = (x_w, y_w, z_w)$ and the corresponding in the camera reference 3D coordinate point $X_{camera} = (x_c, y_c, z_c)$ is defined based on the extrinsic parameters (rotation R, translation T). When dealing with a stereo rig, the camera projection matrices can be expressed as follows:

$$X_{left} = P_{left} \cdot X_{world} = R_{left} \cdot X_{world} + T_{left} \tag{4.1}$$

$$X_{right} = P_{right} \cdot X_{world} = R_{right} \cdot X_{world} + T_{right} \tag{4.2}$$

The task is not only to relate the two image planes in order to associate corresponding points but also to define the relations the image planes and the physical world. For this purpose, we employed Bouguet's algorithm for stereo rectification as presented by Trucco and Verri [45]. This method uses relates the left camera's image plane to the right camera's image plane by applying appropriate rotational and translational transformations forming the projection matrix of the left image plane in accordance with the one of the right camera's:

$$P_{left} = R \cdot P_{right} + T \tag{4.3}$$

with R and T being the rotation matrix and translation vector respectively relating the two planes formed as follows:

$$R = R_{right} \cdot R_{left}^T \text{ and } T = T_{left} - R^T \cdot T_{right} \tag{4.4}$$

where R and T essentially rectify the coordinates of the right camera X_{right} to those of the left camera as X_{Rect_right}. Furthermore, although this transformations lead to coplanar image planes, row alignment is not achieved. To do so a rotation matrix R_{rect} consisting of three epipolar unit vectors with mutual orthogoniallity is computed and applied to the left projection matrix moving the left camera's epipole to infinity as well as aligning the epipolar lines horizontally leading to row alignment of the image planes ([42]). The row alignment of the two cameras is achieved by setting the rotation matrices as:

$$R_{left} = R \cdot R_{rect} \text{ and } R_{right} = R \tag{4.5}$$

The resulting image planes are now aligned and can be searched for point correspondences. By rectifying the image planes we have succeeded in the epipolar lines in the two images to be parallel with the x-axis leading to the corresponding points in the two images being as close to each other as possible, with any point disparities being in the x-axis. The reason that no exact matching is achieved can be attributed to the fact that the application of arbitrary 2D projective transformations can lead to the image being distorted which essentially means that finding the pair of transformations that relates the two images introduces also distortion ([42]).

Essentially, based on these linear relations, we can derive the physical point coordinates to whom each image point is back projected, through a process known as triangulation (as presented in Chapter 2). This process utilizes the image plane point $x_{im} = (x, y)$ and 3D world point $X_w = (x_w, y_w, z_w)$ relations (eq.(3.28)) and parallel projectivity relation to form a system of linear equations based on the cross product:

$$x_{im} \times P \cdot X_w = 0$$

Thus, we end up with a system of linear equations of the form $B \cdot X_w = 0$, where B is a matrix of the following form:

$$B = \begin{bmatrix} x \cdot p^{3T} - p^{1T} \\ y \cdot p^{3T} - p^{2T} \\ x^{'} \cdot p^{'3T} - p^{'1T} \\ y^{'} \cdot p^{'3T} - p^{'2T} \end{bmatrix}$$

with $x_{im} = (x, y)\,, x^{'}_{im} = \left(x^{'}, y^{'}\right)$ the point correspondences in each image, and $P = \left[p^{1T}\ p^{2T}\ p^{3T}\right], P^{'} = \left[p^{'1T}\ p^{'2T}\ p^{'3T}\right]$ the projection matrices.

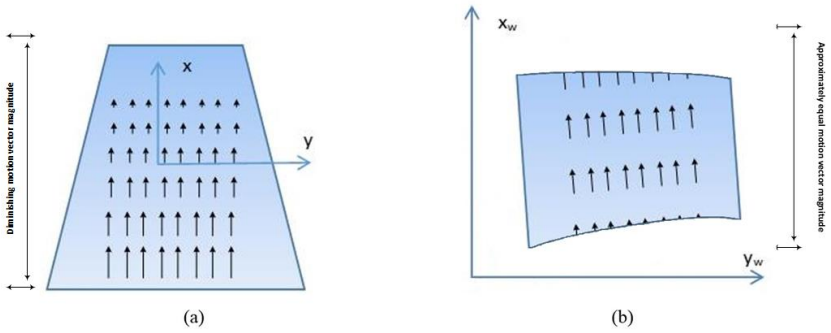


Figure 4.2: (a) The distorted scene representation as recorded from the camera, (b) the undistorted scene representation after the orthorectification process.

In the case of no noise then there will be an exact solution for X_w. However, in the opposite case the image coordinates measurement is inexact and thus, we will have another solution for $B \cdot X_w$ expect zero. To solve the system we apply the constraint that the norm $\|X_w\| = 1$. Essentially, the problem now becomes:

Minimize $\|B \cdot X_w\|$ *under the constraint* $\|X_w\| = 1$: The solution of the linear system and thus, the derivation of the 3D world point is achieved through the singular value decomposition (SVD) of the matrix B using at least 4 known correspondences. The matrix B is factorized in the form of UDV^T, with U and V being orthogonal matrices and D being a diagonal matrix with non-negative entries. The 3D world point is the unit singular vector that corresponds to the smallest singular value of the matrix B(*total least squares*), which is equivalent to the (unit) eigenvector of B^TB with the least eigenvalue, under the assumptions that B^TB is invertible and B is an mxn matrix, with rank n and $m > n$ ([42]).

Nevertheless, in many real-world applications including traffic monitoring, surveillance, human motion and river flow, the motion is still in two dimensions in the world coordinate system, with the third dimension of height being of minor importance. Focusing on such applications, we can limit the back-projection to only two dimensions, which is exactly the case of mapping the perspective onto the projective mapping (see figure4.2), which performs a distance and motion scaling from the camera plane to the 2D world plane parallel to the observed surface.

4.1.3 Disparity and Depth map estimation

The image planes relation performed in the previous step results in the points of each plane to be associated in the form of a translation in x-axis. Thus, we can now define the disparity map d that contains the distance between corresponding points in the two views.

The computation of the disparity map can be performed using similar approaches used for the computation of the optical flow of an image. In our case we employed a variational approach developed by J.Rally [62], due to the fact that it produces denser and more accurate disparity maps compared to a block matching-based disparity estimation method.

What does the disparity map offer?
The disparity map allows us to compute the depth map of the scene by back-projection of each 2D image plane point into its corresponding point position in the 3D world coordinate system. This process, known as triangulation (see figure), utilizes image-point correspondences, as defined by the disparity map, in the image pair of a stereo layout to derive estimates of the Homography matrices, essentially the projection matrices P, which map the 3D world point to its corresponding 2D image point, eq.(3.28).

The derivation of the point correspondences in the two image planes is performed using the disparity map d, which indicates the point relation between the two planes. Specifically, as mentioned, the rectification process led to the relation of the two planes in the form of a simple translation in the x-axis meaning that a point $x_left = (x, y)$ in the left camera image plane is located in the point $x_{right} = (x^{'}, y) = (x + d_i, y)$ in the right camera image plane, with d_i being the disparity for the given point, as shown in figure 4.3.

Finding the point correspondences and through inversing the transformations applied in the rectification process we can find the corresponding coordinates of the points in the original unrectified planes which can now be associated with 3D world point through the mapping relations presented in Chapter 3, in the Stereo Vision subsection, which show that the 3D point estimation process leads up to system of linear equations.

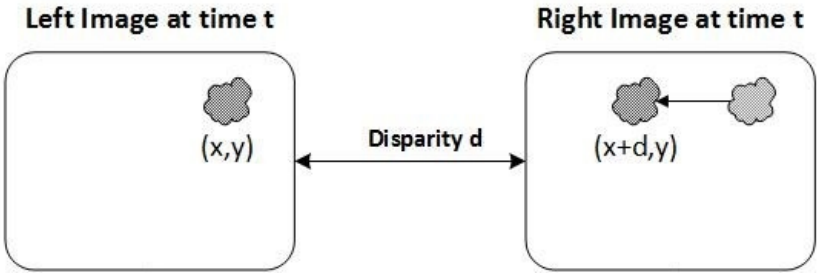


Figure 4.3: The relation of corresponding points in the two views after the plane rectification process.

How will depth information be used?

The additional depth information is used as means of validating the velocity estimate produced by the monitoring system. This is done by using the depth information to define a region of examination in which the velocity estimate is validated by examining the time for which an image pattern (usually a natural tracer, e.g. leaf or foam) takes to move through a region whose length is defined by the depth difference of two objects used as indicators/markers present at the viewed scene. More on this subject will be presented in the final sections presenting our monitoring framework.

4.2 Optical flow Estimation

The next step of the estimation process is the computation of the 2D image velocity field. For this task we have developed a probabilistic method for the computation of the optical flow field of a fluid based on the methodology introduced by Chang et.al. [1]. The approach presented in the following subsections is based on our previous work [2] on probabilistic optical flow estimation schemes.

Probabilistic approaches, as presented in Chapter 3, assume that the motion vector (displacement) of a pixel can be considered as a random variable following a distribution model. Thus, we can formulate a probabilistic inference model which associates to each possible position in the candidate neighborhood in the next frame a probability of displacement. Such formulations allows the employment of Bayesian inference schemes that can predict the motion field. However, in a probabilistic formulation the model's accuracy depends upon on the a' priori assumption about the data distribution characteristics and priors which with their turn are highly affected by the amount of the data that is available. Fewer data will result in an over fitting case where the outliers will

reduce the accuracy of the estimate instead of increasing it. To solve this problem one solution, according to [1] is to use a number of subsequent frames to increase the data used in the estimation process. In our method we proposed the use of a regularization factor λ that will penalize the outliers' importance and thus, keeping the estimate robust without the use of extra data.

Our proposed approach on this issue, is to take advantage of the stereo layout and incorporate the additional scene information provided by the use of a stereo layout into the displacement coefficient estimation process. As will be presented in the following sections this addition ill result in the reduction of the amount of data used in overall estimation process through the reduction of the size of the interrogation window at the examined frame pair.

4.2.1 Probabilistic Optical flow Formulation

We will start by presenting the theoretical basis of the probabilistic optical flow method presented in Bacharidis [2].

The probabilistic method uses a conditional Bayesian model to estimate the unknown 2D velocity field $u = (\upsilon, \nu)$. The Bayesian formulation is based on the assumption that a pixel's flow vector is actually a random variable described by a probability distribution function. So, the unknown velocity can be formed as a posterior distribution that is estimated based on the Maximum a' posteriori rule which is defined from prior likelihood motion model assumptions:

$$\hat{u} = argmax_{\hat{u}} p\left(\phi|u\right) \cdot p\left(u\right) \tag{4.6}$$

where $p\left(\phi|u\right)$ is the conditional probability describing the observed data given the actual realization of the underlying global flow field and $p\left(u\right)$ is a probability describing a prior knowledge for the motion, usually a Gibbs distribution.

As mentioned in Chapter 3, subsection 3.3.1, the conditional probability describes the observed data given the underlying motion. The key factor in this representation is the appropriate selection of the data representation function ϕ. In our approach the representation function, as Chang [1] suggested, is formulated as the probability of a pixel in the reference frame ending up in a specific position in a fixed neighborhood in the next frame:

$$\phi \approx \frac{A_i}{\sum_j^{D_s} A_j} \tag{4.7}$$

where D_s is the size of the candidate neighborhood in which our pixel can be positioned in the next frame.

The coefficient A_i denotes the transition probability of the examined pixel to a specific position i in the candidate neighborhood. These coefficients are estimated by means of the relation of the image intensities of both frames, reference and next, based on a Spatio-Temporal Autoregressive model:

$$I\left(x,y,t\right)=\sum_{i}^{D_s}A_i\cdot I\left(x+\Delta x_i,y+\Delta y_i,t+\Delta t\right) \tag{4.8}$$

where $I(x,y,t)$ is the intensity of the pixel (x,y) at time t in the reference frame and $I\left(x+\Delta x_i,y+\Delta y_i,t+\Delta t\right)$ the intensities of the pixels$(x+\Delta x_i,y+\Delta y_i)$ at the destination neighborhood at time $t+\Delta t$, which in our approach $\Delta t=1$ since we consider the next frame.

The estimation of the transition probabilities A is performed through a least squares scheme that utilizes the pixel intensity information of the neighborhood, N_s of the pixel (x,y) in the reference frame and the pixel intensity information of the candidate neighborhood, D_s, in the next frame. We essentially end up with a minimization of the cost function:

$$J\left(A\right)=\sum_{s=1}^{N}\left(I\left(x_s,y_s,t\right)-k_s^TA\right)^2 \tag{4.9}$$

with k_s being a vector of size D containing the intensities of pixels belonging to the candidate neighborhood for the pixel (x_s,y_s) in the next frame and N being the number of elements in the spatial neighborhood N_s of pixel (x_s,y_s) in the current reference frame.

In order to decrease the effect of outliers and constrain the estimation we have added a regularization term κ which acts as a penalty term on the estimate on the existence of outliers. Thus the previous cost function is now defined as follows:

$$J\left(A\right)=\frac{\sum_{s=1}^{N}\left(I\left(x_s,y_s,t\right)-k_s^TA\right)^2}{N}+\sum_{j=1}^{D}\kappa^2\cdot A_j \tag{4.10}$$

If we represent the previous equation in a matrix form we end up with a least squares estimation problem of the form $Ax=b$, with A being a matrix containing the transition coefficients, x being a matrix containing the intensities of pixels belonging to the spatial neighborhood N_s centered at the pixel (x_s,y_s) and finally, b being a matrix containing

the intensities of the pixels belonging to the candidate neighborhood D in the frame t+1 where the pixel (x_s, y_s) is expected to be displaced at.

In our case this least squares problem is estimated with respect to A instead of x since the matrix A contains the coefficients to be estimated. The solution of such least squares scheme is given by:

$$A = \left(K^T K + N \cdot \kappa^2 I\right)^{-1} K^T M \tag{4.11}$$

with K being the matrix containing the vectors k_s with the intensities for all the possible transitions for each pixel contained in the spatial neighborhood N_s defined by a central pixel (x_s, y_s), and M being the matrix containing with the intensities of the pixels belonging to spatial neighborhood N_s of the central pixel (x_s, y_s).

4.2.1.1 Local Motion Estimates

At this point we can use the estimated transition probabilities to extract a local motion field for the fluid, by selecting as the final destination position the position assigned to the highest transition probability A_i. This local estimation scheme although being able to detect the main motion trend does not suffice for the accurate estimation of the fluid.

This is due to the dynamic and highly consistent motion nature of the fluid. The motion field of a liquid actually shows a unique velocity vector everywhere, meaning that each pixel should be characterized by a single velocity vector which can be extracted using the information of distribution functions we found. Thus, a general motion vector interpolating model must be defined that can estimate and associate the velocity vectors for all pixels belonging to the fluid.

4.2.1.2 Global Motion Estimates

The extraction a global motion field is achieved through the utilization of the Bayesian inference scheme, as presented in Chapter 3.3, in equation (3.18). Through the MAP formulation we can calculate a posterior distribution for each pixel through which we can choose the appropriate distribution that will give us the velocity vector best describing the pixel's movement. This global motion is obtained by maximizing the posterior distribution which is proportional to the minimization of a cost function consisting of a smoothness function term, defined the prior motion model(which in our case is a Gibbs

distribution) and the data function ϕ:

$$\begin{aligned}\hat{u}_{global} &= argmax_{\hat{u}} p((I|u) \cdot p(u) \\ &\simeq argmax_{\hat{u}} p((I|u) \cdot e^{-\lambda U} \\ &\equiv argmin_{\hat{u}} - ln\left(p((I|u) \cdot + \lambda U\right) \\ &\equiv argmin_{\hat{u}} L\left\{\phi\left(\Delta x_i, \Delta y_i\right)\right\} + \lambda U\end{aligned}$$

with L being denoting a functional term based on the conditional distribution and U being an energy based smoothness factor.

In our approach, as Chang et.al. [1] initially presented, the functional term is a differential function consisting of linear relation of the discrete data function ϕ with a family of 2-D Gaussian distributions, thus defining the new functional term $\bar{\phi}$ as:

$$\bar{\phi}_{x_s,y_s,t}\left(x_s + \Delta x_s, y_s + \Delta y_s\right) = \sum_i \left[\phi_{x_s,y_s,t}\left(x_s + \Delta x_i, y_s + \Delta y_i\right) * h\left(\Delta x_s - \Delta x_i, \Delta y_s - \Delta y_i\right)\right] \tag{4.12}$$

in which $(x_s + \Delta x_i, y_s + \Delta y_i) \in D_s$ and $h(x, y)$ is the family of Gaussian basis functions.

The smoothness factor U is linear combination of the image gradients, that define the image energy:

$$U = \sum_s \left(\|\mathbf{u}_x(x_s, y_s)\|^2 + \|\mathbf{u}_y(x_s, y_s)\|^2\right) \tag{4.13}$$

Finally, the cost function leading to the global velocity estimated, minimized through gradient descend based methods, is defined as follows:

$$CostFunction = -\sum_s log\left(\bar{\phi}_{x_s,y_s,t}\left(x_s + \Delta x_s, y_s + \Delta y_s\right)\right) + \lambda \sum_s \left(\|\mathbf{u}_x(x_s, y_s)\|^2 + \|\mathbf{u}_y(x_s, y_s)\|^2\right) \tag{4.14}$$

where $\mathbf{u}_x(x_s, y_s) = \mathbf{u}(x_s + 1, y) - \mathbf{u}(x_s, y_s)$, $\mathbf{u}_y(x_s, y_s) = \mathbf{u}(x_s, y + 1) - \mathbf{u}(x_s, y_s)$ and $\mathbf{u}$ is the global velocity vector.

4.2.2 Stereoscopic Data Utilization

As shown in the previous paragraphs, the estimation of the local displacement probabilities A_i for each pixel position in the candidate neighborhood is an inference problem leading to a least squares solving scheme which is high dependent on the amount of data used. One way to increase the estimation accuracy and at the same time reduce the

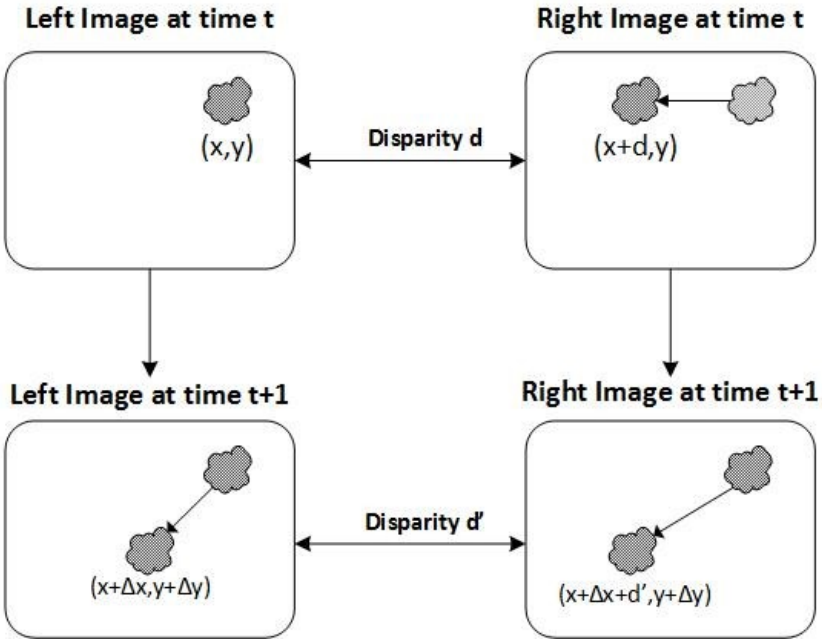


Figure 4.4: Relation between the stereo image pairs and their optical flow estimates based on the disparity map.

amount of data used is to utilize the extra information provided by the bi-channel formulation of the stereo scheme. The pixel intensities of the two camera pairs, that have been set up in a parallel layout or have been rectified when a convergent layout is used, are related based on the disparity d as follows:

$$I_{left}(x, y, t) = I_{RectRight}(x + d, y, t) \tag{4.15}$$

For the subsequent frames, where a displacement for the pixel(x, y) occurs the relation between the paired images for the stereo layout is defined based on both the displacement$(\Delta x, \Delta y)$ as well as the disparity estimate (see figure 4.4):

$$I_{left}(x + \Delta x, y + \Delta y, t + 1) = I_{RectRight}(x + d' + \Delta x, y + \Delta y, t + 1) \tag{4.16}$$

with $d' = d + D_t$ indicating the disparity of the shifted pixel, which is essentially a change in the disparity value of the previous frame at a constant D_t.

For each pixel intensity at each camera based on the STAR model as shown in eq. 4.8 we have the relation:

$$I_K(x, y, t) = \sum_{i=1}^{D} A_i \cdot I_K(x + \Delta x_i, y + \Delta y_i, t + 1) \tag{4.17}$$

where K=left, right denoting the intensity of the image for each camera.

Combining equations 4.16 and 4.17 allows us to express equation 4.15 in the following form:

$$4.15 \Rightarrow \sum_{i=1}^{D} A_{left,i} \cdot I_{left}\left(x+\Delta x_i, y+\Delta y_i, t+1\right) = \sum_{j=1}^{D} A_{right,j} \cdot I_{RectRight}\left(x+d'+\Delta x_j, y+\Delta y_j, t+1\right) \tag{4.18}$$

If we compare equation 4.18 in parts and with consideration that the intensities of each image of the camera pair at time t+1 are connected based on equation 4.16 we can conclude that that given the disparity estimate the coefficient depicting the translation to a certain position must have the same value in both the images in the image pair, i.e. $A_{left,i} = A_{right,j}$, with $i = j$.

As mention the coefficient estimation for the case of a single camera, in a matrix formation solving the least square estimation scheme of the form $Ax = b$ with respect to A instead of x. For the case of a stereoscopic layout the matrices A, x, b are now defined as:

$$x = \begin{pmatrix} I_l\left(x_1+\Delta x_1, y_1+\Delta y_1, t'\right) & \cdots & I_l\left(x_1+\Delta x_D, y_1+\Delta y_D, t'\right) & 0 & \cdots & 0 \\ 0 & \cdots & 0 & I_{rr}\left(x_1+\Delta x_1+d', y_1+\Delta y_1, t'\right) & \cdots & I_{rr}\left(x_1+\Delta x_D+d', y_1+\Delta y_D, t'\right) \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ I_l\left(x_N+\Delta x_1, y_N+\Delta y_1, t'\right) & \cdots & I_l\left(x_N+\Delta x_D, y_N+\Delta y_D, t'\right) & 0 & \cdots & 0 \\ 0 & \cdots & 0 & I_{rr}\left(x_N+\Delta x_1+d', y_N+\Delta y_1, t'\right) & \cdots & I_{rr}\left(x_N+\Delta x_D+d', y_N+\Delta y_D, t'\right) \end{pmatrix}$$

$$A = \begin{bmatrix} A_{left,1} \\ \vdots \\ A_{left,D} \\ A_{right,1} \\ \vdots \\ A_{right,D} \end{bmatrix}$$

and

$$b = \begin{bmatrix} I_l\left(x_1, y_1, t\right) \\ \vdots \\ I_l\left(x_N, y_N, t\right) \\ I_{rr}\left(x_1+d, y_1, t\right) \\ \vdots \\ I_{rr}\left(x_N, y_N, t\right) \end{bmatrix}$$

with A being a $2Nx2D$ matrix, x being a $2Dx1$ vector, b being a $2Nx1$ vector, N being the size of the spatial neighborhood N_s, D the size of the destination neighborhood,

$I_l = I_{left}$ the left image intensity value, $I_{rr} = I_{RectRight}$ the rectified right image intensity value and $t' = t + 1$ the time-stamp indicating the next frame.

Given the assumption that the transition coefficients (A_{left} and A_{right}) for each translation in each image pair must have the same value we can express the coefficient vector A as:

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} A_{left,1} \\ \vdots \\ A_{left,D} \end{bmatrix} = [I : I]^T \cdot A_{common} \tag{4.19}$$

What we essentially did is that we expressed the coefficient matrix A in form $A = C \cdot A_{common}$, where C is a $2DxD$ matrix and A_{common} being a $Dx1$ vector. This formulation allows us to estimate the transition coefficients, removing the duplicates since we have shown that $A_{left,i} = A_{right,j}$, with $i = j$.

By doing this we have doubled the information used to approximate the transition probabilities, modeling the estimate considering the illumination variation encountered between the stereo image pair. This formulation enables us to use smaller windows sizes to reach the same accuracy compared to the case of a single camera layout.

This formulation allows us to estimate a common set of transition coefficients for the two cameras since $A_{left,i} = A_{right,j}$, with $i = j$, utilizing matched intensity information from both cameras as to increase estimation accuracy. Essentially, in this form we double the intensity values used to approximate the transition probabilities for each point and model the estimate by also considering the illumination variation encountered between the stereo image pair. This formulation enables the use of smaller windows sizes to reach the same accuracy compared to the case of a single camera layout. A possible drawback of introducing stereoscopy in the estimation of displacement distribution pertains to the introduction of an error factor associated with the stereo camera model and the computation of an accurate disparity map estimate that provides the correct relation between the pixel points of the stereo image pair. The error sensitivity can be reduced

by constraining the estimated model with the addition of a regularization parameter λ, which controls both the modeling and computation errors for the original ill-posed inversion problem.

Finally, the estimation of the transition coefficients is again performed through the same least squares formulation shown in equation 4.11. Moreover, a global motion field is estimated as shown in the previous subsection through a Maximum a' posteriori (MAP) formulation using the new transition coefficients A_i.

4.3 Velocity Computation in World Coordinate System

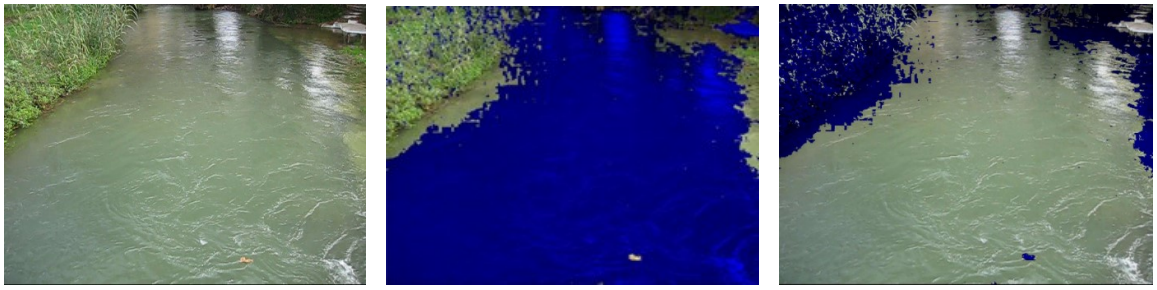
Up to this point we have computed the optical flow field of the scene using the full intensity information from the stereo camera system, but all the computations pertain to the common 2D reference system used by the camera planes. Since we have also derived the disparity map for the two cameras in parallel concatenation, as well as the general concatenation scheme, we also have the means of transforming the 2D back to the 3D world space. Thus, in this section we derive the mapping of the motion vector field from the 2D reference system of cameras to the 3D world coordinate system.

4.3.1 River Isolation and Main trend Extraction

Prior the mapping between the coordinate systems and the extraction the 3D real world velocity of the river, we need to exclude the unwanted information that exists in order to reduce the computational cost as well as to increase the estimation accuracy.

This step involves the isolation of the river from the unwanted terrain, i.e. river banks, rock formations within the water. Such data despite being redundant can also reduce the estimation accuracy since cases such as dispersion phenomena of the water hitting the river banks or the rock formations inside the flow, produce motion vectors that do not depict the river's main trend of motion and thus, are not useful for the estimation of the velocity of the flow. For the case of the river banks although the vegetation along the river may show motion due to the wind or the illumination variance, yet do not provide any information for the flow of the river and thus, must be ignored.

A quick way to deal with unwanted scene information such as the vegetation in the river banks as well as rock formations inside the river flow is to apply a segmentation approach on the region of interest. This will separate the river from the vegetation. To



(a) (b) (c)

Figure 4.5: River isolation via segmentation, (a) Initial image of Koiliaris river, (b) Class containing vegetation, rock formations, leaves and (c) Class containing the pixels belonging to the river.

do so, we transformed the river image into the HSV color space and performed a K-means classification task on the hue and saturation channels of the image, separating it into two classes (see figure 4.5), one containing the vegetation, rock formations, leaves flowing in the water etc. and the other containing the water pixels. In order to distinguish which of these clusters contains the water pixels we used an evaluation metric. The group class that has the biggest average value, computed by the associated pixels, contains the water pixels.

Isolating the river reduces the unwanted flow vectors from the vegetation and the diffusion effects in the river banks and rock formations in the flow. However, this does not solve the surface flow vectors of the river flow that are produced by the wind or the illumination variance. The conventional measuring systems form a tunneling effect in which the water flows to its natural motion direction, without being disturbed by wind changes in the surface of the flow.

Since we cannot constrain the river flow conditions, we need to define a means of distinguishing the main motion of the flow from the motion vectors caused by external forces, such as wind. As a solution we can apply a classification method on the motion vectors of the estimated flow field, in order to isolate and identify the underlying motion trends in the river flow. Motion trend classification will allow the extraction of a main trend of motion for the river flow, that will allow us to increase the velocity estimation accuracy since the affect of redundant motions such vortical motion patterns will be reduced.

How many motion classes? and supervised or unsupervised classifiers?

The biggest questions to every classification problem is first the number of classes selection and secondly, whether to use a supervised or an unsupervised classification method.

Starting from the first question the motion trends can be grouped into 8 motion directions (North, South, East, West and the four subsequent motion directions). Since there is always a hint for the main motion direction of the river we can easily mark which of these motion directions are considered valid for the main trend of motion, e.g. the river flows to the North so we accept North, North-East and North-West motion directions, as shown in figure .

As far as the supervised/unsupervised debate we compared unsupervised (K-NN, K-means, Expectation Maximization (EM)) and supervised (Naive Bayes) classifiers. As will be presented in the Result section, the supervised classifiers presented the best results compared to the unsupervised ones, an expected result since in the case of unsupervised learning the distance and error metrics tested (Euclidean Distance, Cosine similarity) lead to the utilization of a sole discriminant factor (magnitude or direction) in the vector matching process.

4.3.2 River Surface Velocity Computation

The computation of the image based motion vector is performed by computing the pixel displacement in subsequent frames and dividing them by the time interval between the subsequent frames. In order to extract the corresponding motion vector in the 3D space, we need to relate each 2D pixel position with its 3D point correspondence. The stereo layout allows us to derive such mapping through the process of *trianguation*. This process, as presented in *2.5.7.1*, produces a mapping between the two coordinate systems given the fact that an appropriate point correspondence exists between the points of the two image planes.

Thus, the 3D motion vector of a point can now be expressed as the displacement between the two estimated 3D position of each point based on the relative positions of pixel-based correspondent, divided by the time interval Δt:

$$u_{real} = \frac{\|\hat{X_{world,new}} - \hat{X_{world,old}}\|}{\Delta t} \tag{4.20}$$

Since the images used have been orthorectified meaning that it has constant scale in which features are represented in their 'true' positions only two directions of estimated 3-D world point are used the x and y direction. This is due to the fact that the image has been reformed as though every point were viewed simultaneously from directly above.

The velocities computed correspond to the instantaneous velocities of points on the river's surface. The main trend in the river velocity can be considered as the average of several instantaneous velocities in the world system, i.e. for M relevant points of interest:

$$u_{river} = \frac{\sum_{i=1}^{M} u_{real,i}}{M} \tag{4.21}$$

In order for this velocity estimate to be considered as the river's surface velocity we need to validate the assumption that this velocity will almost be constant over the area of examination. Of course, this assumption will only show validity in cases of rivers where the flow is unobstructed and is not accelerated due to the change of the ground's inclination. Notice that the use of a world coordinate system and the projective mapping in the expression of the motion permits the direct averaging of vectors estimated at different points of the river, since they are devoid of perspective effects depending on the viewing depth

4.4 Velocity Estimate Validation

To validate the estimated velocity we can make use the existence of physical particles in the flow, such as leaves or flowing wood parts. Such validation may be performed under the notion that the computed leaf velocity should follow the main trend of the river motion or the fluid velocity on the surface of the river. When using natural particles with physical dimensions, i.e. objects such as leaves, we cannot guarantee that the influence of external forces, such as wind or turbulence, will not dominate the particle's motion forcing the particle to follow a different motion from the typical fluid field. Nevertheless, under the assumption of minimal external forces, the validation process can be performed as follows.

1. **Particle Selection and Detection:**Select the particle to be tracked. The selection can be either user assisted, i.e. the user selects the particle he considers the most appropriate manually in the image, or an automated particle identification can be performed. For the latter we can make use of the segmentation method used in

previous sections. We can observe in figure 4.5 that the particles flowing in the river, case e.g. leaves, have been assigned to the non- water class. We can perform a supervised segmentation process on this class moving to the YCbCr color space, aiming at a specified value range of the channels Cb and Cr that corresponds to the particle's color values, brown color in the leaf's case. The identified particle is then stored as a reference pattern.

2. **Examination Region Formulation and Velocity Validation:** In order to simplify and constrain the validation process, we can define a region of examination through which the particle's motion will examined.

   The validation is performed by searching the specified particles after a number N of frames, under the assumption that the estimated 3D velocity estimate remains constant over the river length. The constant velocity is used to match the particle's 3D position at the initial time stamp and the estimated position based on motion estimation. Afterwards, by back-projecting the estimated position on the camera plane(s) we examine the particle's presence within the specified region of the frame(s).

In the following subsections details on the validation process will be presented as well as how the region-dependent parameters used in the validation process are defined.

4.4.1 Validation Process

Instead of starting from the region formulation we will begin our methodology presentation by first defining the overall validation process and then explaining the way that the region-dependent parameters are defined in the next *Region Formulation* subsection.

The velocity validation can be performed in two ways:

1. **Time specified particle search:** in which given the estimated velocity and a region with known real world dimensions we estimate the time required for the particle to travel the region. The search for the particle in the corresponding estimated frame exiting the region validating the estimate.

2. **Position specified particle search:** in which given the estimated velocity, the initial particle position and a predefined time interval, then we can estimate the final position of the particle. Essentially, using the estimated 2D motion vectors

and the projection relation between the points of the 2D image plane and the one of the 3D world we can back project the final 2D position (due to the motion displacement) to its corresponding 3D position.

In our validation scheme we adopt the first approach estimating the time required for the particle to travel from its current computed 3D position (defined by its centroid) to the position of the first indicator (entry to region of interest) as well as the second indicator (exit from region of interest) given the estimated average 3D velocity:

$$t_{mov} = \frac{X_{w,indFin} - X_{w,indStart}}{u_p} \tag{4.22}$$

in which $X_{w,indFin} = (x_w, y_w, z_w)$ corresponds to the stereo-based estimated 3D coordinates of the centroid belonging to the indicator object denoting the end of the region, $X_{w,indStart} = (x_w, y_w, z_w)$ corresponds to the stereo-based estimated 3D coordinates of the centroid belonging to the indicator object denoting the start of the examination region, and finally, u_p corresponds to the particle's velocity which assumed that it follows the river's motion pattern can be assumed equivalent to the river's motion, $u_p \approx u_{river}$.

Having found the required time we can then move to the future frames and search for the particle near the 'end of region' indicator(see figure 4.6). The number of frames forwards that we need to move in order to search the particle can be easily found given the camera's frame rate:

$$N_{frames} = t_{mov} \cdot \text{frame rate} \tag{4.23}$$

Then by simply moving at the frame = reference frame + $(N_{frames} \pm 1 frame)$ we can search the particle. The particle will be searched in an interrogation window centered at the centroid of the 'end of region' indicator whose width will equal to the image's width dimension and the length will be predefined, as shown in figure 4.6.

The verification of the particle at this position validates both the estimation accuracy as well as the velocity-constancy assumption. The most important assumption behind the velocity estimation method and its validation is the velocity constancy in its 3D form, which is devoid of perspective effects associated with the 2D mapping. In real-world applications, the 3D velocity constancy assumption can be justified better than in the 2D case, where velocity needs to be adjusted with time and scene depth as to accommodate the perspective effects.

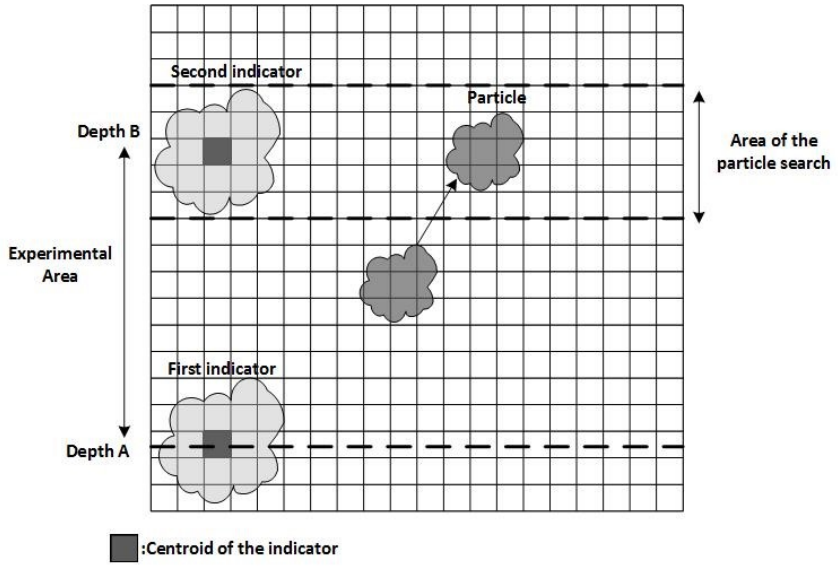


Figure 4.6: Formulation of the search area for the particle in each image plane by defining a window with a horizontal size equal to the image's width and vertical size the size of the particle ± M pixels up and down. The window is formulated around the centroid of the second indicator.

4.4.2 Region Formulation

As presented in the previous subsection in order to simplify, constrain and reduce the computational burden an examination region is defined based on two objects serving as region start and finish markers(indicators). The stereo-driven 3D world coordinate estimates of their centroids allow the computation of the region's length used in equation 4.22 which computes the time required for the particle to move through the examination region.

Two key questions arise in this process: *(a) How do we select the markers/indicators?, and, (b) How do we detect the particle in the frames?*

Starting from the first question the indicator objects, are assumed to be always present in the viewed scene, in static positions. The objects are manually selected by a user, by defining a rectangular window around the object, in the image depicting the viewed scene. The 2D centroid of the pixels belonging to the object is defined and its corresponding estimated (via the stereo layout) 3D coordinates are extracted. The length of the region

can then be defined as by simply taking the depth difference (z-axis coordinate estimate) of the centroids of the two object-indicators.

As for the second question the reference particle pattern will be searched within this region using the SURF feature extraction and identification methodology presented by Bay et.al. [63]. As previously mentioned the presence of a particle or a characteristic particle portion within this region can be considered for the accuracy assessment of the estimated instantaneous velocity of the river.

To further increase the estimation confidence we can also back-project to the expected particle position in the two cameras. In such case, the particle's position will be different on each image plane of each camera defined by the direct mapping from 3D to 2D via the corresponding matrix projection. Then, by taking advantage of the disparity map we can relate the pixel positions and the motion vector estimates between the two image planes. We can use this relation to search the particle in the two camera planes thus, increasing the tracking validity of the particle.

Chapter 5

Results

In this chapter we present the results of our work. The result section will be divided in four sections presenting and commenting upon the following:

- **The 2-D stereoscopic probabilistic optical flow estimation method:** Illustrating the improvement in accuracy as well as data reduction compared to the use of single camera probabilistic method.

- **The comparison of the proposed stereo-based technique with other relevant methods:** Test and compare the proposed optical flow method with other approaches utilized in the fluid flow extraction case in order to illustrate the efficiency of the proposed method to capture local motion trends retaining the both magnitude and directional accuracy.

- **The ability of summarization of motion trends:**Perform clustering on the underlying motion directions.

- **The overall developments and stages of 3D motion estimation:** Presentation of the results for each stage of the proposed imaged-based river monitoring system.

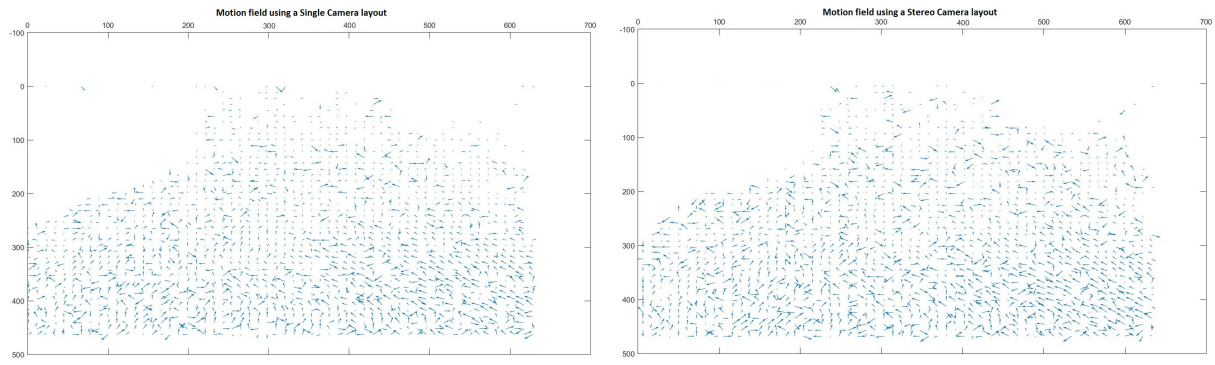
Finally, an additional section is added presenting a visual environment(a graphic user interface) which incorporates the aforementioned stages of image-based river monitoring system into a user-friendly tool.

5.1 Improvement of 2D motion estimation using stereo data

Apart from the 3D river velocity estimation framework, in this thesis we have also presented a new stereo based probabilistic optical flow estimation method. In the following paragraphs we will present how the incorporation of the stereo-driven information in the optical flow estimation process, enhances the estimation accuracy but also allows the reduction of the amount of data used in the estimation process. During this evaluation stage we compare the stereo-based approach with a single camera optical flow estimation methodology presented in my undergraduate diploma thesis [2]. The two approaches are compared, with real as well as synthetic datasets, on the estimation accuracy, the amount of data used and the post-processing capabilities that each method presents.

5.1.1 Accuracy using Real Data

As far as the optical flow estimation accuracy is concerned, we observed small deviations the optical flow field estimates of both approaches. Specifically, based on the experimental results taken in Koiliaris River, show that approximately 21.7% of the main motion vectors change and about 9.85% of the total motion vectors between the two optical flow derivation schemes. Both approaches succeed on retaining the main motion trend information as shown in figure 5.1. However, their main difference lies on their trade-off between the estimation accuracy and the amount of data used to achieve it. Based on the amount of data used, the stereo case requires a smaller interrogation window size $\sim$ 38% to reach an estimation accuracy with less than 10% deviation from the single camera approach. This is due to the incorporation of the data from the second camera which acts as an enhancement for the estimation process. In terms of 3D real world velocity estimate deviation between each approach this $\sim$ 22% main motion vector deviation results in the single-based approach to produce an estimate of $0.4201 m/secs$ compared to the $0.3864 m/secs$ of the stereo approach with the Conventional equipment producing $0.3993 m/secs$.



(a) (b)

Figure 5.1: (a)The estimated motion field using a single camera and stereo camera layout,and, (b)Estimated motion field using the on-field stereo layout scheme.

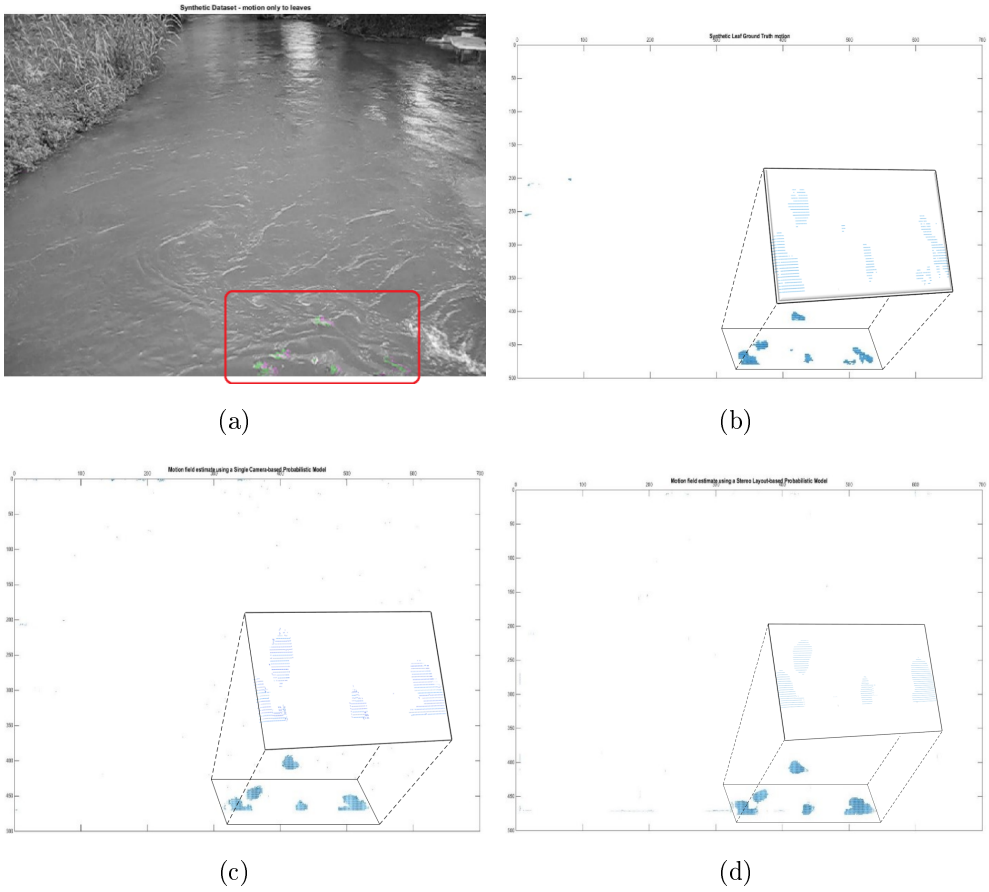
5.1.2 Accuracy using Synthetic Dataset

To further back these observations we have tested both of these approaches with synthetic datasets using ground truth motion fields. The synthetic datasets where created based on the monitored scene. Motion fields where applied only to the leaves present at the flow. This serves a double cause, since despite the motion estimation accuracy testing it also indicates the ability of our approach to adapt to particle tracking, through the use of color-based segmentation approaches.

This dataset was generated for both single and stereo camera layouts. The motion formulations employed in the synthetic dataset consist of linear translations in the each axis separately and diagonal translations ranging from 0 to 4 pixels. For the stereo case the disparity map d maximum translation in the x-axis (due to parallel layout assumption) is of approximately 5 pixels. Figure 5.2 presents one of the generated synthetic datasets and the ground truth motion field. The selection of the translational motion patterns instead of rotational or spiral motions was based on the facts that the main trend of motion in the river flow which is followed ideally by all the tracked particles tends to follow ideally (assuming no rock formations exist within the river which may produce rotational or spiral motions due to diffuse effects) a linear translational motion pattern.

The single and stereo based approaches are compared both on the optical flow estimation accuracy as well as the on the amount of data (window size) required. The optical flow field estimates of both the single and stereo optical flow estimation approach are compared to this ground truth optical field. The estimation accuracy assessment is performed based on the Angular (AE) and Endpoint (EE) error metrics ([64]). Table 5.1

shows the average observed errors using the same window size (13 x 13) for each of the motion cases.



(a) (b)

(c) (d)

Figure 5.2: (a) The synthetic reconstructed river scene with motion added on the leaves, (green) indicates the initial position and (magenta) the shifted position (a translational motion to the right direction), (b) the corresponding ground truth motion field, (c) the estimated optical flow field using the single camera probabilistic optical flow method, and (d) the estimated optical flow field using the stereo camera probabilistic optical flow method.

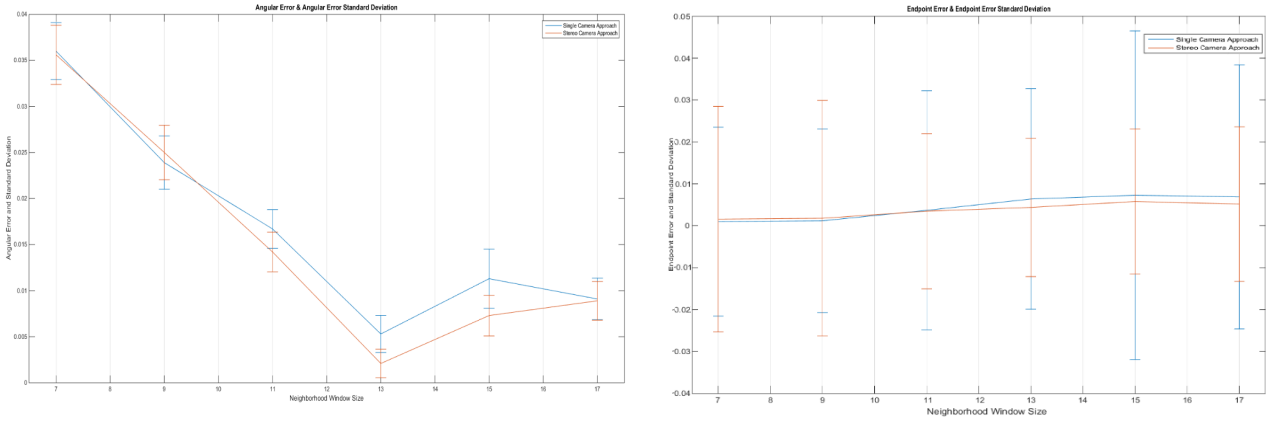
Method	Angular Error(AE)	Endpoint Error(EE)	Stdev AE	Stdev EE
Single Camera	0.0162	0.0066	0.0037	0.0283
Stereo Rig	0.0155	0.0039	0.0034	0.0209

Table 5.1: Endpoint and Angular errors between the Stereo and the Single camera Probabilistic optical flow estimation approaches using for the stereo case a neighborhood size 20% smaller than the one used for the single camera optical flow probabilistic model.

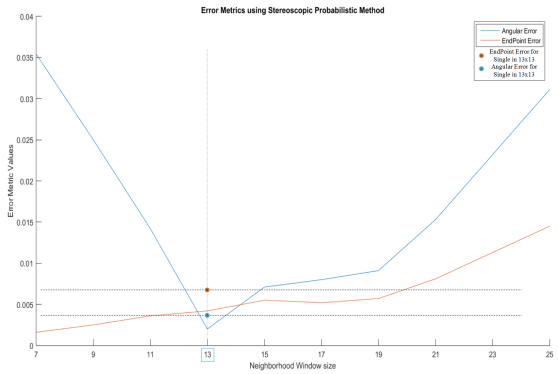
We observe that both approaches lead to high accurate estimates with the stereo-based approach retaining more information about the magnitude information compared to the single-based method with a small loss in the directional information. The resulted errors show that a mean error Endpoint error of 0.2 pixels and an Angular error of lesser than 1 degrees is produced by the application of a Bayesian inference optical flow estimation technique. These results indicate a very good estimation accuracy in the estimation for both the directional and the amplitude information of the motion field. The suitability of the Bayesian inference approach in the optical flow estimation for fluids flow fields has been proved in Bacharidis [2].

However, as mentioned, the most important benefit of the stereo method is the data reduction during the estimation process. In our test cases we have observed that the use of the stereo based probabilistic optical flow estimation method can lead to approximately 15.82% dataset size reduction, without presenting loss in accuracy, compared to the use of the single camera probabilistic optical flow estimation method. Figure 5.3 presents the relation between the window block size variation (window size range $[7-17]$) and the estimation accuracy for both the single-based as well as the stereo-based approaches. In subfigures 5.3.a and 5.3.b we present the error variance for different window sizes between the two approaches, whereas subfigure 5.3.c presents the fluctuations of the error metrics opposed to the optimal window size for the single camera approach (solid points, window size 13x13).

Commenting on the results, we can observe that the stereo-based approach outperforms the single camera-based method. The stereo based method achieves the same or better accuracy using smaller window sizes, thus reducing both the data requirement as well as the computation time. Moreover, we can also observe that the standard deviation of the error metrics is also reduced despite the size increase of the interrogation region indicating that the produced flow field is more compound. As for the estimation accuracy of the presented stereo-based method, the Angular and Endpoint error metrics range below 1 degrees and 0.01 pixels, respectively, implying good estimation efficiency.



(a) (b)



(c)

Figure 5.3: Average Angular and Endpoint Errors and their standard deviations using various window sizes in the examination neighborhood formulation for the case of the stereo-based probabilistic optical flow estimation method. (a) AE and StD of AE for both methods, (b) EE and StD of EE for both methods, (c)AE, EE for the proposed stereo-based approach compared to the optimal neighborhood window size found for the single camera probabilistic optical flow estimation method(13 x 13 window- marked with blue, bullet points indicate the corresponding AE and EE).

5.2 Efficiency of Motion Estimation in Capturing Local trends

In this section we evaluate the proposed Stereo-based Probabilistic Optical flow estimation method on three basic types of fluid motion that can be encountered in both laboratory and real world experiments. The motion models used to formulate the experimental dataset consist of: (a) a simple linear motion, (b) a vortex pair with a rotational motion pattern, and (c) a membrane-like motion arrangement. These motion patterns were synthetically formulated by applying the appropriate motion function to a set of moving particles. The generation process follows the synthetic motion generation scheme based on Kalman filter, which is implemented in the PIVLab tool (Thielicke and Stamhuis [65]; Thielicke [66]). The flow fields are seeded with particles of 3-pixel diameters, with a size variation of ± 0.5 pixels, having a displacement range from 0 to 4 pixels. Synthetic images are used as "ground-truth" patterns, since the flow motion pattern and the particle seeding are well defined. In particular, ground truth motion fields are derived from the motion equations used to create the synthetic images.

The algorithms compared are (a) a Gaussian weighted Lucas-Kanade method (Baker and Matthews [12]), (b) the classic Horn and Schunck method (1981), (c) the Single Camera Probabilistic approach (Bacharidis [2]) using normal pdfs and (d) a local Particle Image Velocimetry tool (OpenPIV) (Thielicke and Stamhuis [65]) that utilizes a cross-correlation algorithm with shifting windows to accommodate for large displacements and reduce the effect of out- of- plane motion(Taylor et.al. [67]; Adrian [56]) and finally, (e) the proposed algorithm..

In every experimental case we have used a macro block formulation of 16 x 16 pixels, leading to a sparser and easier to comprehend flow field. This was performed so as to first, match the first 2 methods and the proposed method with the block based formulated methods (c) and (d). The aim behind the selection of these methods is to illustrate the difference between the fluid directed methods from the classical optical flow approaches as well as the difference between local and global philosophies in the estimated optical flow field.

The evaluation metrics used to assess the accuracy of each method are the Angular Error (AE) and the Endpoint Error (EE) (Baker et.al. [64]). The Angular Error is determined by the dot product of the motion vectors divided by the product of the motion-vector lengths, followed by the inverse cosine transformation; it depicts accuracy

of the directional information on the estimated motion field. On the other hand, the Endpoint Error is defined as the absolute squared error between the estimated and the ground truth motion vectors; it primarily depicts the magnitude accuracy for the estimated motion field. We evaluate each method's accuracy based on the average and standard deviation (STD) of these error metrics over the entire image, which are presented in the following tables. These metrics allows us to have a qualitative inspection on the estimated motion vectors.

Linear Translation Motion Case:
The linear motion case implies small deviations in amplitude in the rage of $[0, 3]$ pixels. In order to allow for estimating such small deviations, the smoothing factors in all methods is not as severe as it should be in uniform motion. The results of each method are shown in figure 5.4 and the corresponding Error metrics in Table 5.2.

Method	Angular Error	Endpoint Error	Stde AE	Stde EE
Proposed	1.3882	0.1164	0.3383	0.0441
Single Cam Probabilistic	1.5339	0.1732	3.8183	0.1693
Weighted Lucas-Kande	8.2863	0.9624	4.6570	0.1090
Horn-Schunk	2.8214	0.5508	0.8821	0.1052
Particle-Based(OpenPIV)	1.8335	0.2338	3.8095	0.2470

Table 5.2: Evaluation metric results for the linear shift motion case.

From the results we observe that the proposed method as well as the single camera probabilistic method and the local particle-based method surpass in accuracy (both in direction and magnitude) the differential approaches. More specifically, we observe in Table 5.2 that the proposed method produces average angular and endpoint errors of $\sim$ 1.4 degrees and 0.1164 pixels, respectively, implying good estimation efficiency. Moreover, the standard deviation between the estimated vectors is small both in magnitude and direction implying a coherent and accurate estimate. Both the stereo (proposed) and single probabilistic (Figures 5.4 (a),(b)) methods preserve the discriminant information of the region and thus result in high fidelity estimates, but with the stereo (proposed method) showing reduced errors compared to the single camera one.

Differential approaches do not perform well in recovering magnitude information and result in significant underestimation of the motion effects. The weighted Lucas-Kanade

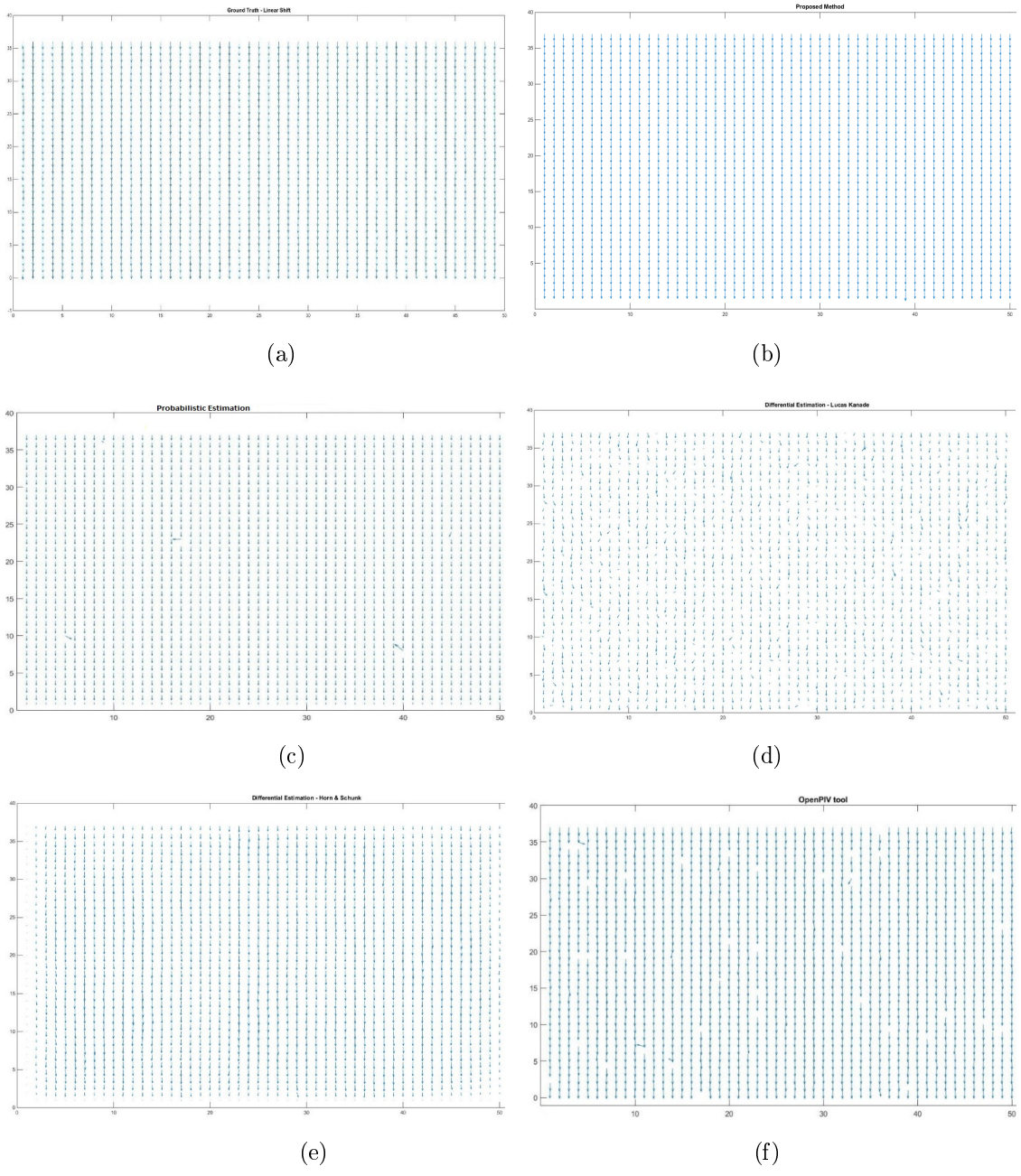


Figure 5.4: Linear shift case; (a) Ground Truth motion field, (b) Stereo-Based Probabilistic method(**Proposed**), (c) Single Camera Probabilistic approach, (d) Weighted Lucas - Kanade, (e) Horn-Schunck and (f) OpenPIV tool.

method(local one) shows the worst performance in both amplitude and direction estimation due to the locality of the implementation scheme forming a trade-off between uniform and locally varying estimate. Horn-Schunck's method(global one) allows better preservation of the directional information of this flow field, yielding low angular errors with a small deviation factor (Figure 5.4(d)). However, from Table 5.2 we observe that Horn and Schunck's method is inferior to the recovery of magnitude information compared to the probabilistic and particle based methods due to the fact that the smoothness induced by the regularization factor used, over-weights the motion amplitude at each region. Finally, as previously mentioned the local particle-based method achieves high estimation accuracy coming third to the comparison process.

Vortical Motion Case

For the second test of a rotational motion pattern with two vortexes, the estimated motion field varies smoothly with a diminishing effect in magnitude while approaching the vortex's center.The results of each method are shown in figure 5.5 and the corresponding Error metrics in Table 5.3.

Method	Angular Error	Endpoint Error	Stde AE	Stde EE
Proposed	8.4615	0.4080	3.2832	0.1627
Single Cam Probabilistic	10.3751	0.4530	4.0843	0.1297
Weighted Lucas-Kande	13.0720	0.4999	6.9035	0.3148
Horn-Schunk	10.0056	0.5989	3.2550	0.3860
Particle-Based(OpenPIV)	5.4748	0.2172	5.0570	0.8076

Table 5.3: Evaluation metric results for the vortex pair of the rotational motion case.

In this test case, the proposed approach(Figure 5.5(a)) ranks second in performance behind the particle-based methods showing an average Angular Error of $\sim$ 9 degrees and average Endpoint Error of $\sim$ 0.5 pixels showing an error decrease of $\sim$ 1.5 degrees and $\sim$ 0.05 pixels from the single camera Probabilistic approach(Figure 5.5(b)). From the top performing method of this scenario the particle-based method(Figure 5.5(e)) the proposed approach shows an error increase of 3.5 degrees. Nevertheless, the standard deviation of both the stereo(*proposed)* and single camera probabilistic schemes is smaller implying that the estimates are more compound. This is a useful algorithmic attribute in more random cases where the estimated displacement is expected to show large deviations affected by noise.

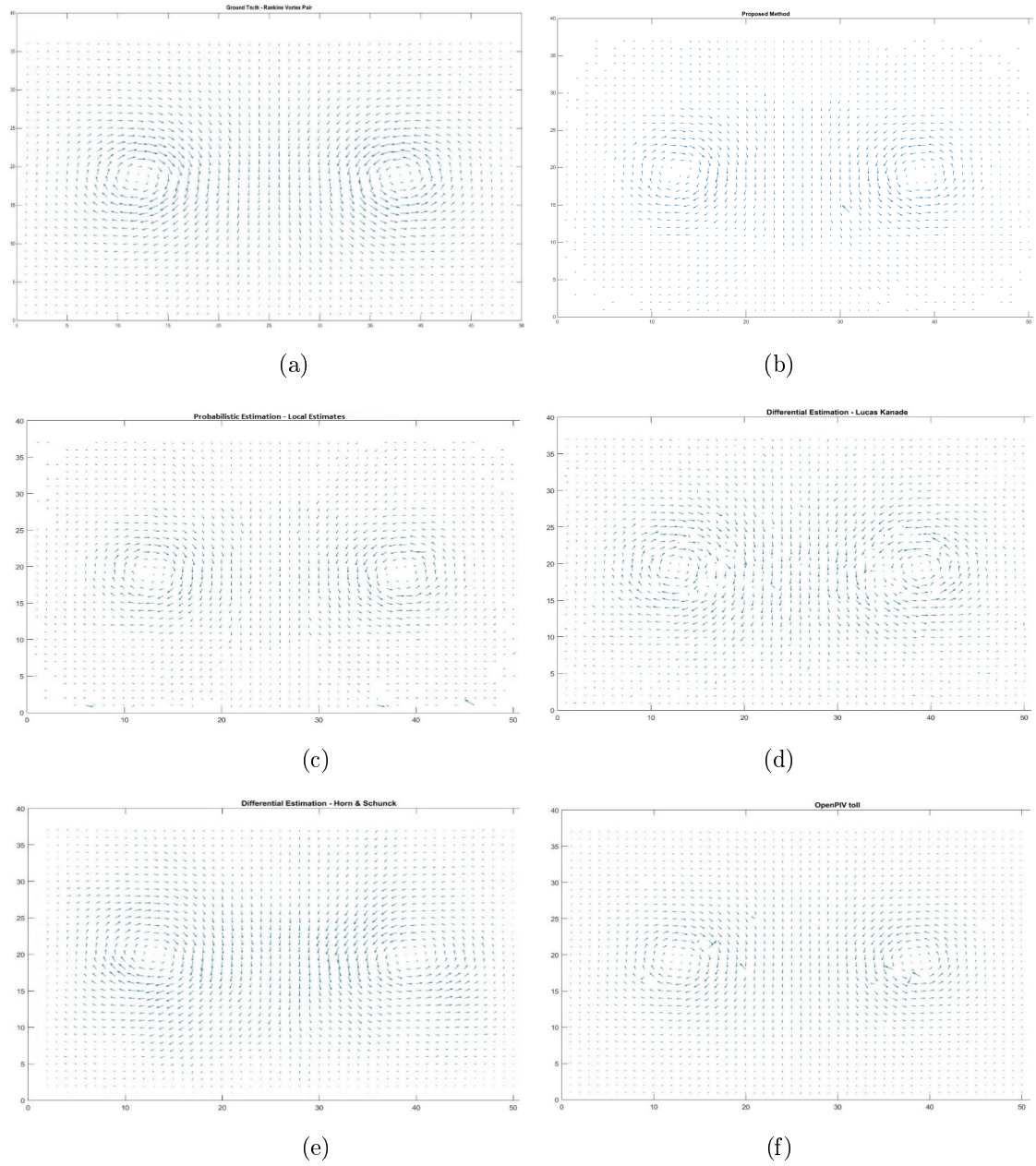


Figure 5.5: Vortex pair case; (a)Ground truth motion field, (b) Stereo-Based Probabilistic method(**Proposed**), (c) Single Camera Probabilistic approach, (d) Weighted Lucas - Kanade,(e) Horn - Schunck and (f) OpenPIV tool.

The differential approaches show increased estimation accuracy compared to the previous problem formulation. Horn- Schunck's method (Figure 5.5(d)) manages to retain

the directional information of the flow, ranking second in the angular error estimation, however, it fails to accurately estimate the magnitude of the motion vectors yielding the highest endpoint error due to the regularization term which in the estimation process although providing a smoother and coherent flow field, yet fails to retain the motion detail of each region. The locally weighted Lucas Kanade scheme on the other hand, manages to counterbalance the smoothness loss with an accurate magnitude estimate. The directional information loss shows an angular error of almost 13 degrees and the magnitude information loss an endpoint error of 0.5 pixels.

Membrane Case

For the final test case of a membrane, the motion pattern is similar to a wave-like motion exhibiting large motion deviations within each region.The results of each method are shown in figure 5.6 and the corresponding Error metrics in Table 5.4.

Method	Angular Error	Endpoint Error	Stde AE	Stde EE
Proposed	5.1023	0.1946	1.8143	0.1244
Single Cam Probabilistic	10.7231	0.2874	2.0170	0.1648
Weighted Lucas-Kande	12.5725	0.2675	1.8950	0.1097
Horn Schunk	20.9866	0.5029	5.0894	0.2584
Particle-Based(OpenPIV)	5.7395	0.1428	2.1470	0.2308

Table 5.4: Evaluation metric results for the membrane motion case.

For the last test case, the proposed algorithm(Figure 5.6(a)) achieves the highest accuracy considering the direction information with an angular error of 5.1 degrees. As far as the magnitude inference task the proposed algorithm ranks second behind the particle based approach(Figure 5.6(b)) with a difference of ~ 0.051 pixels. Considering the results of the probabilistic class of methods (proposed and the single camera based one),the probabilistic methods show a more compound motion field (Figure 5.6 (a),(b)) in terms of magnitude deviation, but have about twice as high angular error compared to the particle based approach. Comparing the two methods we can observe again that the stereo-based(*proposed*) surpasses in accuracy the single camera probabilistic approach, showing reduced estimation deviations both in the directional as well as the magnitude aspect of the motion field.

As far as the Differential approaches they rank last in this test case with the Lucas-Kanade local scheme being slightly better (Figure 5.6(c)), but with errors similar to the global probabilistic method. However, we observe that the resulting motion field is denser and the deviation in the angular error is smaller compared to the global probabilistic scheme. Horn and Schuck's scheme (Figure 5.6(d)) provides the worst estimate due to the fact that its global scheme along with the smoothness constraint enforce strict coherence between regions in the form of smoothness, leading to a continuously diminishing motion estimate towards the center of each dense membrane .

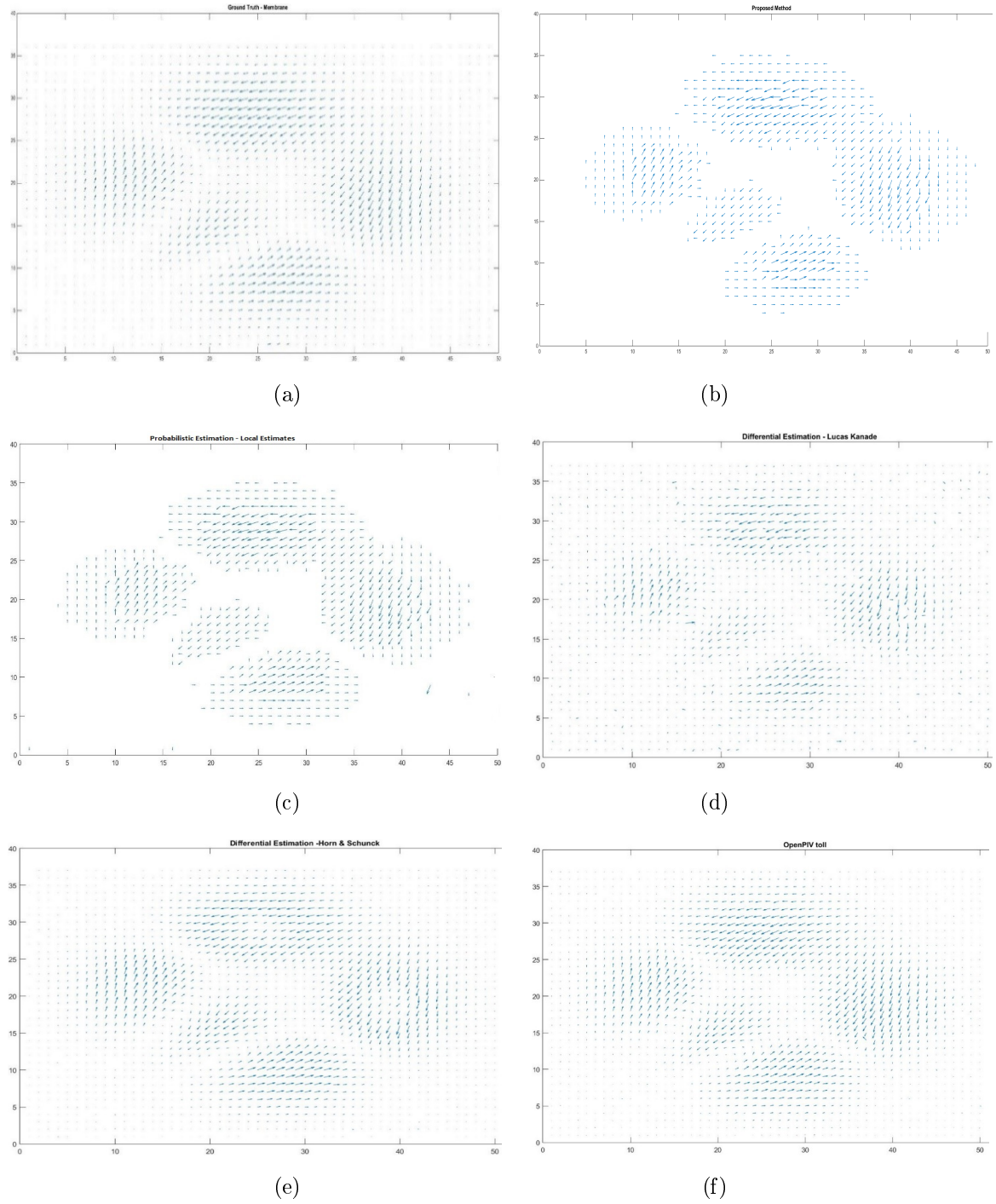


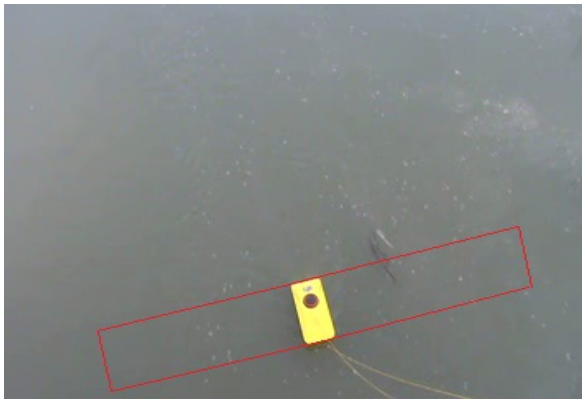
Figure 5.6: Membrane case; (a) Ground truth motion field, (b) Stereo-Based Probabilistic method(**Proposed**), (c) Single Camera Probabilistic approach, (d) Weighted Lucas - Kanade, (e) Horn-Schunck and (f) OpenPIV tool.

5.3 Summarization and Clustering of Motion Directions

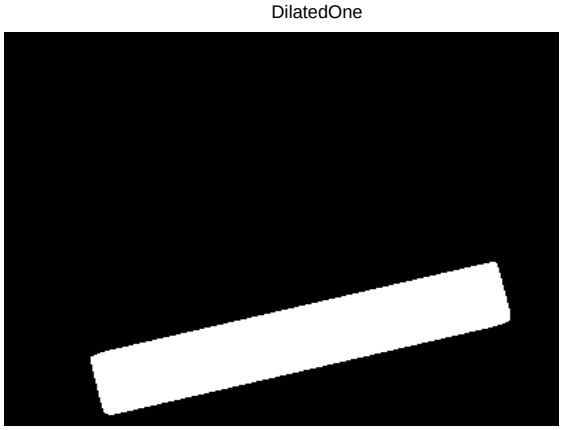
In this section of the Results Chapter we present the comparison results between the selection of a supervised and an unsupervised classification method on the extraction of the main motion trend. We have compared the classifiers both on the main trend of motion selection as well as the real-world velocity estimate and the optimal number of classes. The desired main motion trend should consist of upstream motion vectors with none or small deviation to the left and the right. The examined classifiers are: (a)K-Nearest Neighbors, (b)K-means with Euclidean distance and Cosine Similarity, (c) Expectation Minimization and (d) Naive Bayes classifier. The test case examined in the following experiments regards a monitoring session in Axelloos River. The reference real world surface velocity, determined using a Doppler-based device(*Q-liner*), was found to be $0.6567 m/sec$. In this monitoring session the interrogation area was formed around the Doppler device (Q-liner) whose dimensions were used to approximate the dimensions of the region, as shown in figure 5.7. The underlying motions patterns were found based on the best performing classifier, that will be presented in the following paragraphs.

Selection of number of classes:
Figure 5.8 shows the deviation between the classification result between different classification methods as well as the deviation introduced by the number of classes used. As we can observe in the figure the K-NN classifier although produces a close to the expected velocity estimate, yet it fails to recognize and classify correctly all the wanted motion vectors present in the flow, compared to the Naive Bayes classifier and lacks stability of classification as well as accuracy even with more than 4 classes of motion.

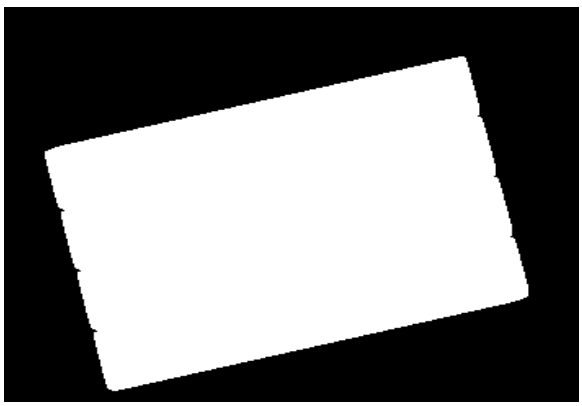
As mentioned the examined motion directions used during the experiments were essentially the four main motion orientations(up, down, right and left) provides the most stable classification results. As we can see in subfigures (c) and (d) of figure 5.8 in both the cases of K-NN and Naive Bayes classifier the use of an extra class leads to misclassifications and thus, a false estimate for the main trend of motion.



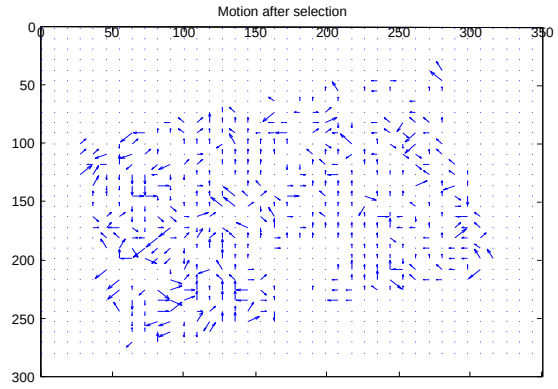
(a)



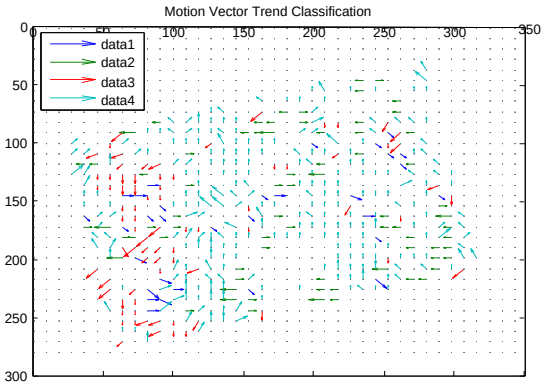
(b)



(c)



(d)



(e)

Figure 5.7: (a) Initial frame(red box denotes the initial dimensionally known work area), (b) Initial work area(white box) based on Q-liner's length dimension, (c) Expanded work area, (d) Optical flow field of the work area, and, (e) Classification result of the underlying motions

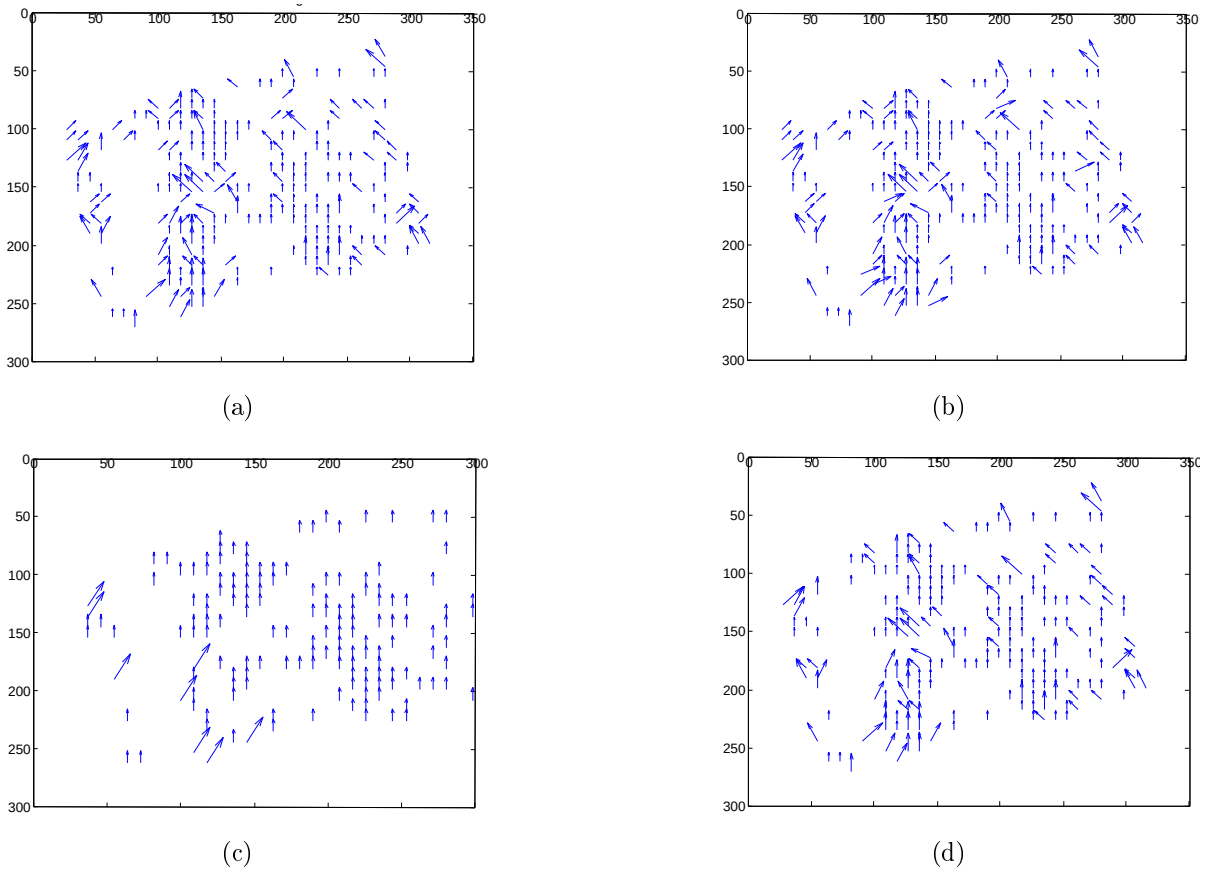


Figure 5.8: Main trend of motion classification deviation between K-NN and Naive Bayes classifiers prior the global optimization step, (a) K-NN with 4 classes, real-world velocity estimate=0.6350 m/sec, (b) Naive Bayes classifier with 4 classes, real-world velocity estimate=0.6377 m/sec, (c) K-NN with 5 classes, real-world velocity=0.7820 m/sec and (d) Naive Bayes with 5 classes, real-world velocity estimate=0.7334 m/sec.

Results based on the real world velocity estimate:

Concerning the estimated velocity magnitude, we compare the classifiers based on the resulting surface velocity estimate, by also examining if they estimated correctly the main motion trend. The classification task was performed 10 consecutive times for each classifier, with the same optical flow field, using the four classes.

The estimated velocities are presented in Tables 5.5 and 5.6. All the classifiers presented a success rate from 5/10 to 10/10 on classifying as a main trend of motion purely upstream motion vectors. The resulted surface velocities presented in Tables I and II are

based on these successful classification results and contain 5 of the 10 resulted values. In the last two lines of the tables are depicted the average deviation the correct classification values as well as the number of correct classifications. As correct classification we define the classification that indicates as a main trend of motion the upstream motion(motion vectors with a upwards or slightly deviated to the left or right motion direction). The deviation between resulted velocity values is due to the initial centroids' values used which resulted in different number of classified motion vectors which combined with the magnitude based metrics used by most of the classifiers yields deviations in the resulted motion. Table 5.5 presents the estimated velocity by using only the local displacement probabilities to derived the motion vectors and Table 5.6 presents the estimated velocities after the global optimization stage of the velocity field, as presented in Chapter 4.

m/sec	K-means (Euc)	K-means (Cos)	K-NN	EM	Naive Bayes
1	0.8418	0.6350	0.7690	0.6198	0.6377
2	0.8893	0.8592	0.7655	0.5962	0.6377
3	0.6350	0.8592	0.5962	0.6259	0.6377
4	0.6350	0.6350	0.8592	0.5920	0.6377
5	0.6350	0.6350	0.7655	0.5328	0.6377
$SucNum$	5/10	6/10	6/10	7/10	10/10
$AvDev$	0.0963	0.0816	0.1325	0.0716	0.0190

Table 5.5: Real-world velocity estimates before global optimization step, Q-liner estimated velocity:0.6567m/sec

m/sec	K-means (Euc)	K-means (Cos)	K-NN	EM	Naive Bayes
1	0.6920	0.7996	0.5318	0.5318	0.6476
2	0.6568	0.6270	0.6728	0.5819	0.6476
3	0.6920	0.6270	0.7521	0.6476	0.6476
4	0.7990	0.7330	0.8055	0.5318	0.6476
5	0.6748	0.8604	0.7521	0.5819	0.6476
$SucNum$	8/10	6/10	6/10	7/10	10/10
$AvDev$	0.0555	0.0853	0.1009	0.0869	0.0091

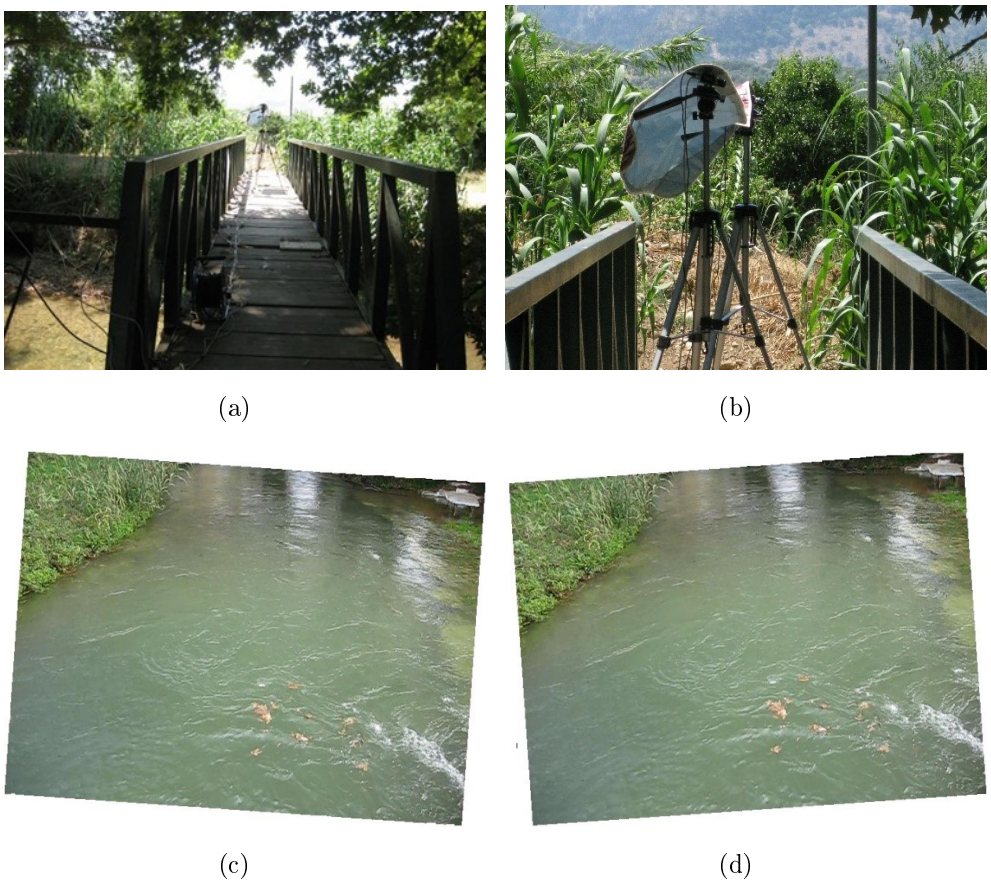
Table 5.6: Real-world velocity estimates before global optimization step, Q-liner estimated velocity:0.6567m/sec

In both cases, based on the fact that Q-liner's estimated velocity is 0.6567 m/sec, we we derive that, prior the global optimization step, the most stable classifier from the unsupervised class is the Expectation Maximization classifier with 7/10 correct identifications and an average deviation from Q-liner's estimation of $0.0716m$ and posterior the global optimization step the most stable classifier is K-means with a deviation of $0.0555m$ from the expected surface velocity. From the supervised class we used the Naive Bayes classifier as a corresponding method. The Naive Bayes classifier presented a better classification accuracy compared to the unsupervised classifiers with a 10/10 correct identifications and an average deviation from the Q-liner's estimation, prior the global optimization step, of $0.0190m$ and posterior the global optimization step with a deviation of $0.0091m$ from the expected surface velocity indicating that a supervised algorithm leads to better accuracy compared to an unsupervised one.

5.4 River Motion Estimation in World Coordinates

We have designed an initial version of the monitoring system that we have previously presented which has been tested in Koiliaris river, Chania, Crete. In our study site, Koiliaris river has a small width ($\approx 8m$) so both parallel and convergent layouts can be implemented. We have selected and placed in the nearby bridge a convergent layout(Figure 5.9(a,b)), consisting of CCD cameras(Prosilica GC1020) to examine how the

convergent layout will perform at a small scale experiment (Figure 5.9). The camera pair is rotated about 12 degrees in the y-axis and ±4 degrees in the x-axis respectively in order to have a panoramic view of the scene. The baseline between the camera centers is $88cm$ and the cameras are rotated about 2 degrees and -3 degrees in the x-axis for the left and right camera respectively. Finally, the camera pair placed at tripods have 5.25 meters distance above the rivers surface. The final system will have the cameras placed at permanent positions on the bridge and with a wireless transmission system in place (in our initial experiment an Ethernet connection to a modem was used).



(a) (b)

(c) (d)

Figure 5.9: Koiliaris River site,(a)The convergent stereo rig placement on top of the bridge, (b)Close up on the stereo rig used, consisting of two CCD cameras, (c)View of the Left camera, (d)View of the Right camera without rectification.

Starting from the stereo layout and the depth estimate, we first calibrated the cameras

by placing a chessboard pattern at the river banks, gaining the information required to relate the physical and the image coordinate systems and thus, producing the 2D to 3D point mapping, information that can be later used to rectify 2D motion field correcting the perspective effect.

The next stage is to extract the 2D motion field from the recorded image data. Figure 5.10 shows the motion field extracted using the proposed Stereoscopic Probabilistic optical flow estimation method. Moreover, using a Naive Bayes classifier(which as will be presented in the following subsection proved to be the best motion pattern classifier) we were able to classify the estimated motion field into 4 motion classes: (a) Pure x-axis translation with increase in x-value, indicating motion to the East, (b) Pure x-axis translation with decrease in x-value, indicating motion to the West, (c) Downwards motion(South, South-East and South-West) and finally, (d)Upwards motion(North, North-East and North-West).

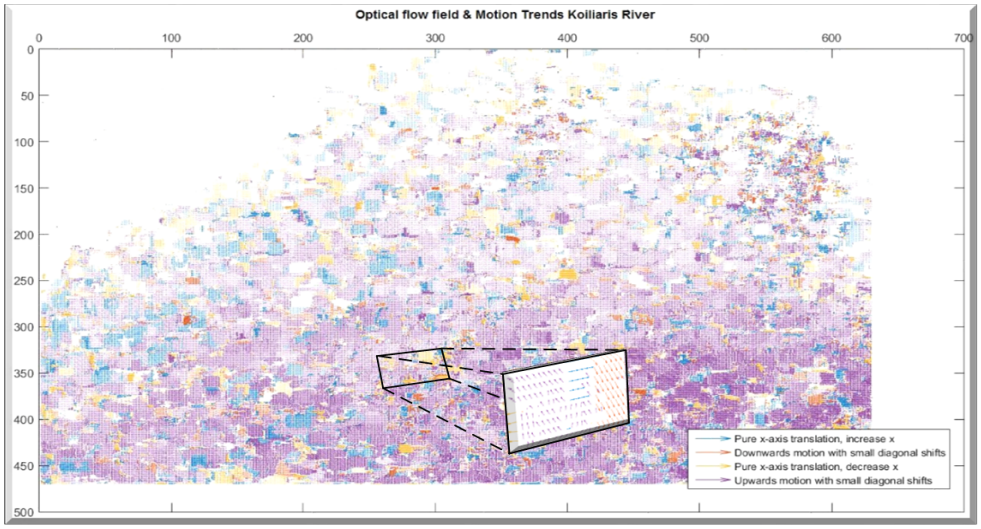
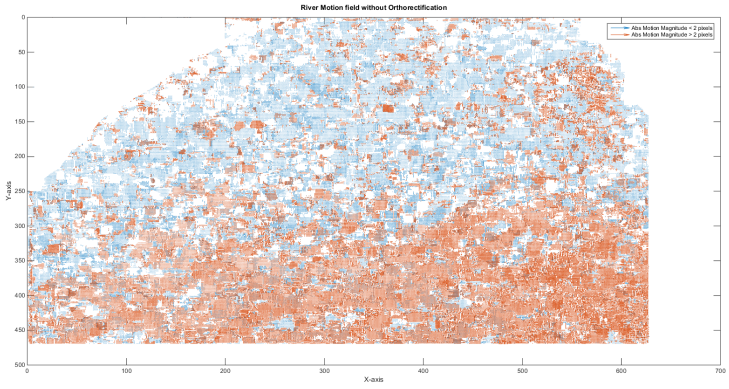


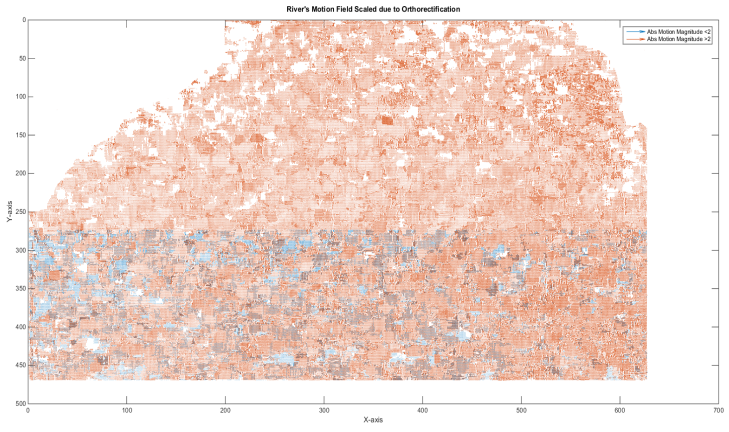
Figure 5.10: An example of the estimated motion fields and the derived main trends using stereo camera layout and stereo-based probabilistic optical flow method.

As can be observed in figure 5.10 the main motion trend is the upwards motion, a result verifying the observable real motion of the river. Prior the motion pattern discrimination and classification we need to isolate the motion vectors belonging to the river from the vegetation, in order to exclude unwanted and erroneous motion vectors from the motion pattern classification process as well as the velocity estimation process, as shown in figure 4.5.

The most crucial aspect of the presented methodology in leading to accurate estimates is the projection correction that this stereo-based approach offers. The perspective affect can be corrected in the stereo based case using the image plane and scene relation resulting in a more accurate flow estimate. This effect manifests itself as in pixels in the far sight of the camera having smaller motion vector magnitudes compared to the pixels in the front of the camera despite the fact that the motion is actually the same. As mentioned existing monitoring systems with single camera layouts required the use of Ground Points with known coordinates in order to undo the perspective effect, a constraint that a stereo layout does not have, thus making this monitoring system more flexible and autonomous compared to its peers.

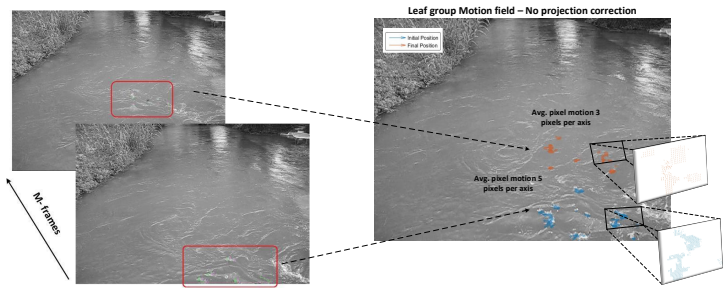


(a)

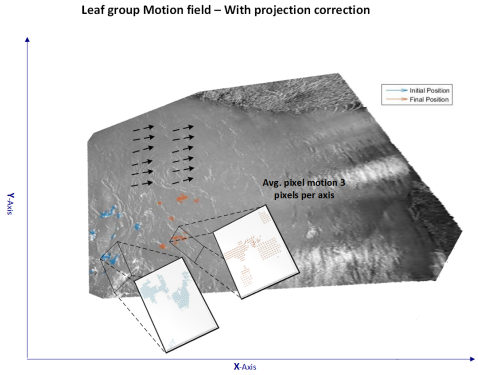


(b)

Figure 5.11: Koiliaris river test case, (a) river's motion field without projection correction placed in image coordinate system and (b) scene and motion field rescaled motion field based on the reconstructed 3D scene coordinates.



(a)



(b)

Figure 5.12: Leaf bunch tracking in Koiliaris river test case, (a) leaf bunch being tracked and corresponding motion field without projection correction placed in image coordinate system and (b) scene and motion field orthorectification based on the reconstructed 3D scene coordinates.

To illustrate this effect we have isolated leaves flowing in the river surface tracking their motion in both viewing cases. The stereo based approach allows the correction of the perspective effect by associating the 3D world point information of the scene with the image plane information and the derived motion field.

Figures 5.11 and 5.12 present the test case and the rescaled optical flow field based on the real world motion as estimated using the relations between the motion changes in each coordinate system. The first indicates the rescaled magnitude of the motion vectors for the whole river flow field whereas the second the rescaled motion vectors for a leaf bunch. The projection correction essentially scales the motion vectors according to the 3D world coordinate system. In the test case presented in Figure 5.12 the 3D motion found in the initial and final leaf positions remains almost unchanged despite the fact

that the image based motion differs approximately 2 pixels.

5.4.0.1 Proposed and Conventional Velocity Estimations

Having found the 2D motion field, derived the transformation relation between the image plane and the 3D world coordinate system, allowing the correction of the perspective effect, we finally reach to the out-most aim of this paper, i.e. the estimation of the 3D surface velocity of the river.

By relating the two coordinate systems, as presented in Chapter4 we can compute the 3D average surface velocity of the river. In the following Table 5.7 we present our velocity estimates on the Koiliaris's surface velocity for 4 monitoring sessions, with measurements taken under almost ideal measuring conditions (no rain or cloudy days). Moreover, we compare the estimated surface velocity with the velocity estimates produced using conventional Hydrological equipment(accelerometers and Doppler based devices) in order to present the estimation deviation. The first two estimates were made the same day at two distinct video recording, whereas the remaining two monitoring sessions where performed in different days.

Test Cases	Conventional Equipment	Proposed Monitoring System	Deviation
1	0.3993 $\pm$ 0.0331 m/sec	0.4124 m/sec	+0.0131 m/sec
2	0.3993 $\pm$ 0.0331 m/sec	0.3864 m/sec	-0.0129 m/sec
3	0.6251 m/sec	0.6009 m/sec	- 0.0242 m/sec
4	0.7143 m/sec	0.6947 m/sec	-0.0196 m/sec

Table 5.7: Table Deviation between the estimated river surface velocities measured with conventional equipment and the proposed image-based method.

As can be observed from the table our approach shows an average deviation of $\pm$0.01745 m/sec between the real river surface velocity measured using conventional equipment (accelerometers and Doppler-based devices) and the velocity estimate derived using the proposed image based method. The velocity estimate using conventional equipment (i.e. accelerometers) is considered to be the average surface velocity $\pm$the variance, as measured throughout the monitoring session. This deviation shows that the proposed approach even at this current prototype stage produces close to actual surface velocity estimates. Even more this deviation can be further reduced through the performance and accuracy optimization of each of the system's components.

5.4.0.2 Velocity Estimate Validation

The final stage of the system, consists of the estimation validation which ensures the accuracy of the estimation. If natural particles, such as leaves, are present within the flow then they are used as validation points for the estimated velocity. The leaves are assumed to follow the river motion, thus having the similar velocities. Assuming an almost constant velocity we simply examine whether the leaf bunch enters an interrogation region, defined by two points or objects in the scene (indicators), after a specific number of frames since the frame that they were first identified.

Following the methodology presented in Chapter 4 we have applied this validation scheme to our experimental scene in Koiliaris River. In our case a leaf formation is automatically isolated through appropriate thresholding in the vegetation class found during the river and vegetation class, its SURF features are being extracted and then the particle is searched in future frames. A simplified example of this process, in our case of study, is shown in figure 5.13, where initially the leaf pattern is isolated and used as a reference pattern. This leaf pattern is then examined in future frames. At each frame each particle is isolated, its SURF features are being extracted and then compared to the reference pattern.

A more elaborate representation of our evaluation scheme, applied in Koiliaris River, is shown in figure 5.14, showing the final implementation of our method applied in one of our monitoring sessions. The red and blue color variations in the figure indicate the leaf motion between consequent frames and were added to highlight this motion. As indicator points/objects we used distinct rock and vegetation formations that were present in the scene forming an interrogation region of approximately 0.59 meters in length, computed by taking the depth estimate difference of the centroids of each indicator(in the test case a bush and rock formations as shown in Figure in points 1 and 2 respectively).

The first leaf encounter within the interrogation region was in frame 128 of our test video. The frame rate of the monitoring camera set was 30fps. The estimated velocity according to our approach was 0.3864 meters/sec. Following our approach we estimated that the leaf bunch will require 86 frames to first exit the interrogation region, i.e. the leaf will require $t = 1.53secs$ to go through the interrogation region. As shown in figure 5.14 the leaf bunch enters the interrogation region in frame 212 and starts to exit (first leaf to exit) the interrogation region at frame 301 showing that our approach had a deviation of $f_{real} - f_{expected} = 301 - 298 = 3$ frames showing that our approach is close to the real

leaf velocity, and thus, proving the validity of the estimated velocity.

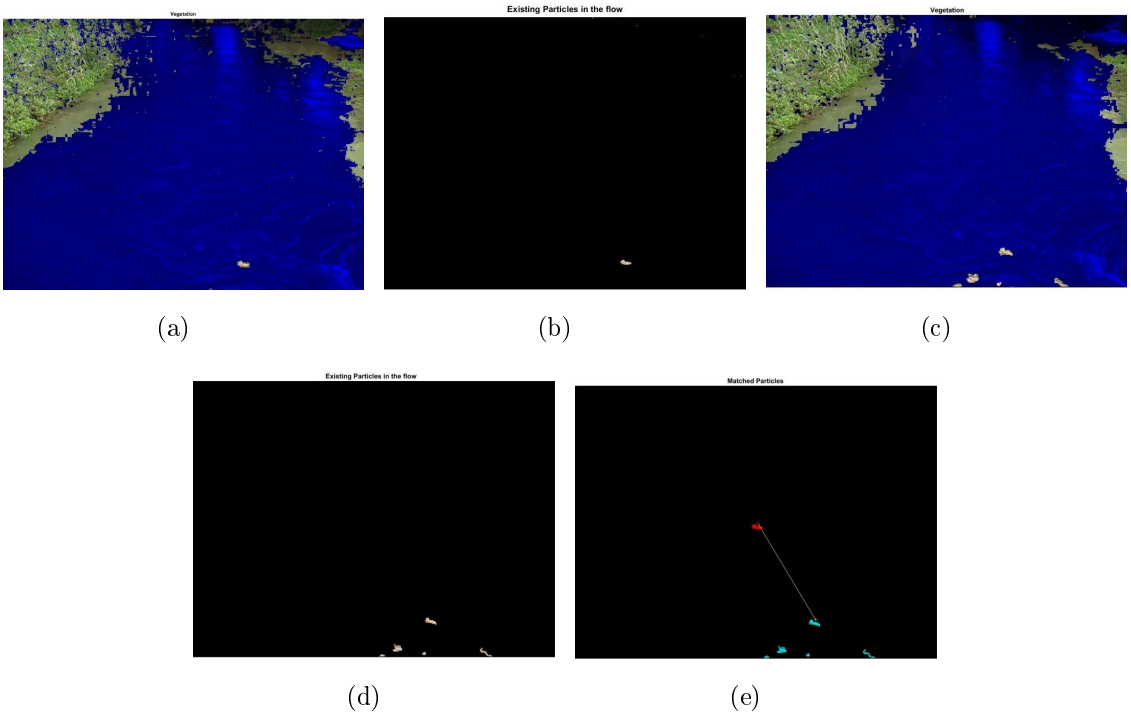
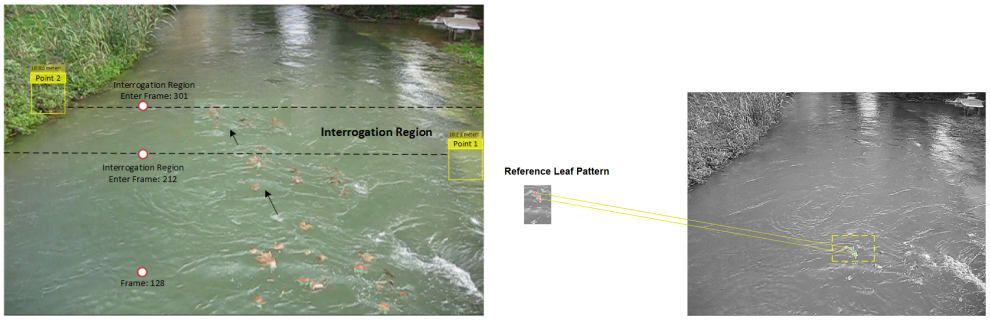


Figure 5.13: Particle isolation and tracking, (a) Segmented scene containing the vegetation and the natural particle being tracked, i.e. the leaf, (b) Isolated particle for the examination process, (c) The vegetation and natural particles present at N = 10 frames forward, (d) Isolated particles present in the current frame and (e) Matching of the reference particle in the current frame to verify that it has cascaded to this expected region at this current frame based on the velocity estimate, with red is the reference particle pattern and blue the particles present in the current frame.



(a) (b)

Figure 5.14: Validation stage in Koiliaris River. Red and blue colors indicate the motion between consequent frames. (a) The leaf bunch formation across the frame series are fused into one unique frame to better represent the methodology stages across the frame series, (b) Identified and tracked leaf formation based on a reference leaf pattern taken from previous frames.

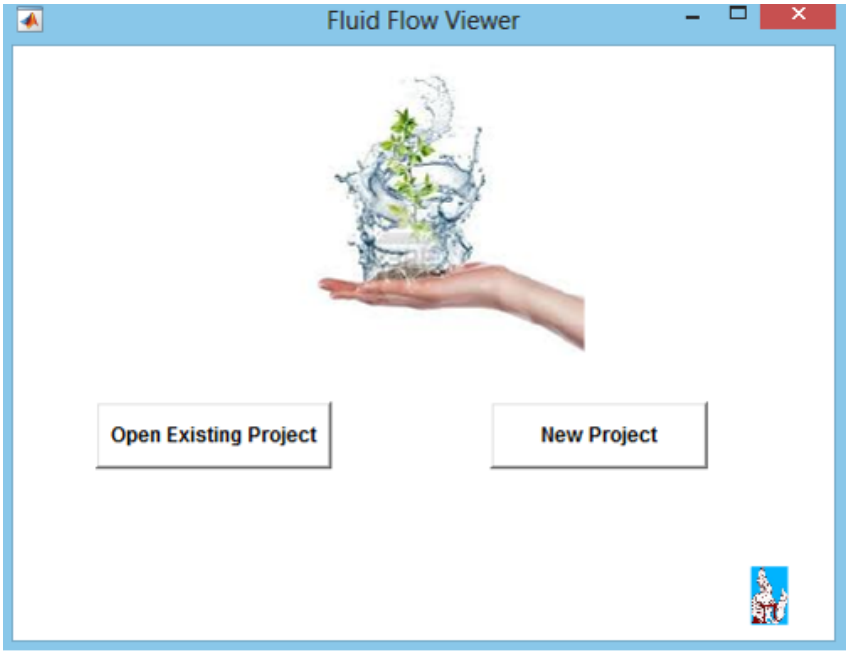


Figure 5.15: Initialization window of Fluid Flow Viewer (v.1.0).

5.5 Visual Environment: Fluid Flow Viewer

Following the work presented in Bacharidis [2], we present a new version of the graphical tool "Fluid Flow Viewer (FFV)" that incorporates all the aforementioned capabilities assisting a user to easily set up an image-based river monitoring system(Figure 5.15). Such tools, can also be very useful to hydrologists studying the properties of the ow and using the information derived for the extra extraction of flow characteristics.

This version of the graphical tool(FFV v.1.0) presents the user with the following possibilities:

- Real time view of each camera and instant single frame and video acquisition.
- On spot stereo rig calibration, plane rectification and depth map extraction.
- Reprojection error display in order to evaluate the depth extraction accuracy.
- Motion field extraction using the presented Stereo-based Probabilistic method.
- Vortex Detection and main motion trend extraction.

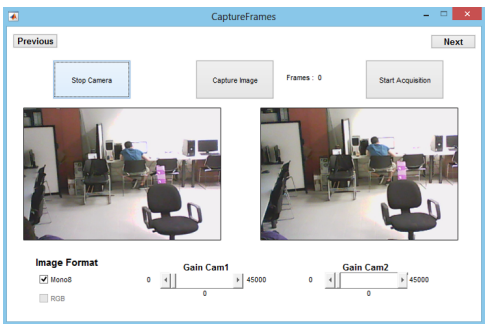
- User- defined selection of the interrogation region by allowing the user to manually select the two indicator/objects, presenting their depth estimates in order to evaluate the accuracy of the depth estimate.
- Real world average surface velocity of the river.

The specific version of the Fluid Flow Viewer was specifically designed in order to allow the user to set up the image monitoring system, acquire images of the flow, calibrate the system and extract on spot measurements of the flow.

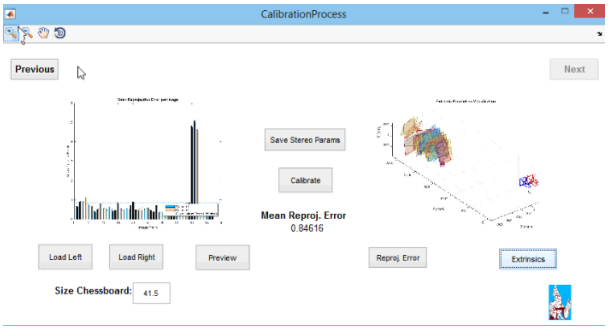
The system's stages involve the following tasks: (a) View of the scene(real time) from each camera allowing the user to appropriately set up the stereo rig. (b) Stereo Rig calibration in order to extract the parameters that allow the relation of the image plane with the real world coordinate systems, (c) Depth Map extraction, (d)Interrogation region formulation through user-defined indicators,(e) 2-D motion field and main trend extraction as well as vortex detection(more details on the methodology used in this stage can be found in Bacharidis [2]), and (e) Real world velocity estimation. Figure 5.16 presents examples of some of the stages.

The steps (d) to (f) can be performed multiple times, allowing the user each time to determine parameters of the Optical flow estimation method as well as interrogation region thus, allowing to essentially determine the accuracy of the final estimate. Moreover, the user is able to save and load the parameters of a river case, i.e. the parameters relating the cameras of the stereo rig and the real world as well as the optical flow algorithm's ones allowing him to perform post-processing on the data.

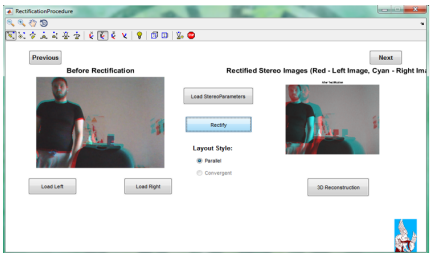
The advantages and novelty of the viewer is that it allows even at this non-optimized prototype on spot measurements with a waiting time between each session of approximately 10 minutes. All the aspects of the monitoring system are controlled from this tool presenting the user the freedom of forming the layout in the most accurate way possible by evaluating the accuracy of stereo rig's relations and depth map extraction. The biggest advantage of this tool is that it provides the system with independence and flexibility to be applied in almost every river case.

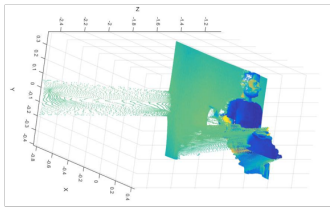


(a)

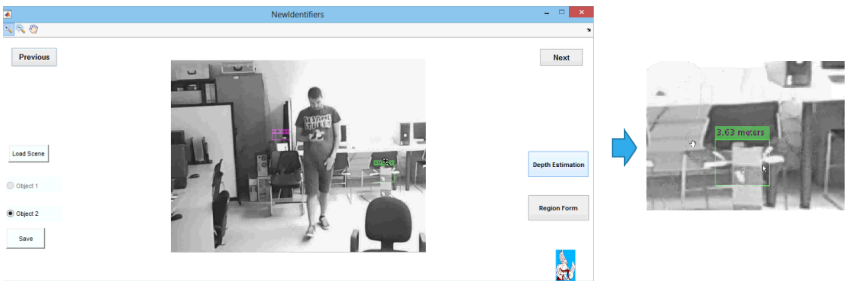


(b)

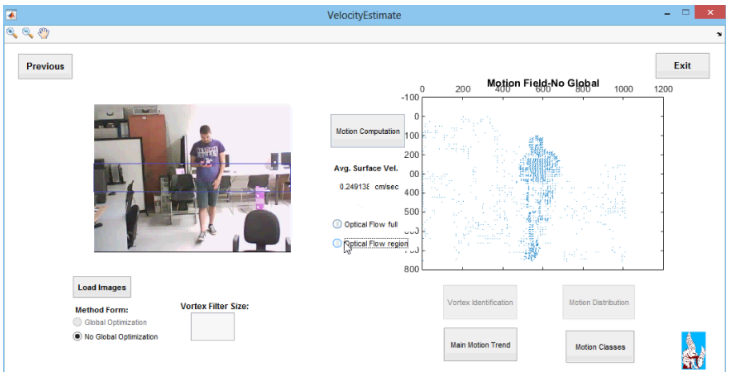




(c)



(d)



(e)

Figure 5.16: Some of the Stages of FFV, (a)Scene view and Frame capture, (b)Camera Calibration and Error Visualization,(c)Depth Map Creation(*image from another session*, (d)Indicator Selection, (e) Region formulation and Motion field

Chapter 6

Conclusion

6.1 Conclusion

This thesis presented a new image-based river monitoring framework that allows the estimation of the real world average surface velocity of the river flow using image data. More specifically, we presented and examined the theoretical basis behind each component of the system and assessed its role in the estimation process. The proposed system shows high estimation accuracy, a deviation of $\pm 0.01745\ m/sec$ from the estimate acquired using conventional equipment, it is autonomous and flexible allowing on-spot measurements at almost any river flow case. Furthermore, compared to its peers the systems is free of the ground points requirement due to the novel use of a stereo camera layout that allows immediate mapping between the image plane and world coordinate systems.

Moreover, this thesis presented a stereoscopic probabilistic optical flow method that utilizes the stereo data as a means of strengthening the displacement probability estimates of a pixel to a candidate position found by the use of Bayesian inference optical flow field estimation scheme. The algorithm was based to a probabilistic method presented in our earliest work(Bacharidis [2]) which for a candidate neighborhood associates each destination position to a transaction likelihood, considered as a sample of a local distribution function. The estimated local distribution functions are combined with a differential basis functions to form a conditional probability of the observed data based on the underlying motion field. By applying the Maximum a Posterior rule, using a prior assumption of the motion model, we can end up with the posterior probability of motion, i.e. the motion field.

The presented algorithm proved to be more accurate compared to particle-based block matching and differential optical flow methods concerning cases of fluid flow motion pattern estimation. The presented method succeed on retaining both magnitude and directional information of the motion field, leading to coherent motion field estimates with angular errors of less than 1 degree and endpoint error of less than 0.2 pixels. Furthermore, compared to its predecessor the addition of the stereo data in the estimation process leads to an increase in estimation accuracy as well as in computational efficiency since in order to achieve the same accuracy the proposed algorithm required almost 15.90% less data in order to estimate the transition coefficients, compared to its single camera predecessor.

6.2 Future Work

6.2.1 Proposed Image-based Monitoring System

As far as the proposed monitoring system, the system will be tested in more river monitoring cases in order to further examine its accuracy and assess the role of each of its parameters in the deviation on the velocity estimate. Moreover, testing of monitoring under various weather conditions, will allow us to assess its performance based on the weather condition factor, since so far the system has only been tested under ideal weather conditions.

Moreover, more elaborate pre-processing and post-processing will be performed on the observed data in order to reduce the affect of noise or extreme illumination variations on the estimated motion field. Techniques such as histogram equalization and motion vector filtering have already applied at the current version of the system, however, the methods used are basic and more problem-focused ones are expected to further increase the accuracy.

6.2.2 Stereo-based Probabilistic Optical flow Method

Since the proposed algorithm is a variation of the one presented in our earliest work, future variations on this that will allow us to further increase its accuracy and computational performance follow the same pattern as the suggestions made for its predecessor(Bacharidis [2]). For example, the use of shifting windows for the candidate

neighoborhood based on an similarity measures will allow the algorithm to detect larger displacements, thus increasing the algorithms accuracy in cases of high-speed fluid flows. Another approach is to change the basis function used to define the conditional probability of motion, i.e. functions whose data representation resembles the examined motion style can be used, allowing a more problem-specific motion field estimation method that increases the accuracy. Finally, concerning the problem of increasing the computational performance of the algorithm one solution is to perform the data computation on Fourier space domain.

References

[1] Chang, J., Edwards, D., Yu, Y.: Statistical estimation of fluid flow fields. In: ECCV workshop on Statistical Methods in Video Processing, Copenhagen. (2002) 91–96 2, 31, 32, 34, 73, 74, 77

[2] Bacharidis, K.: Fluid flow estimation using video data. Diploma thesis, Technical University of Crete, Greece (2014) 2, 3, 22, 73, 74, 90, 93, 95, 116, 117, 119, 120

[3] Chorin A.J. and Marsden J. E.: A Mathematical Introduction to Fluid Mechanics. Springer-Verlag (1990) 16

[4] Cenedese A.: Eulerian and lagrangian velocity measurements by means of image analysis. Journal of Visualization **2**(1) (1999) 73–83 18

[5] Westerweel J., Dabiri D., and Gharib M.: The effect of a discrete window offset on the accuracy of crosscorrelation analysis of digital piv recordings. Exp. Fluids **23** (1997) 20–28 18, 22

[6] Meinhart CD, ST Wereley, JG Santiago: Piv measurements of a microchannel flow. Exp. Fluids **27** (1999) 414–419 18

[7] Hadad T., R.G.: Effects of particle size, concentration and surface coating on turbulent flow properties obtained using piv/ptv. Experimental Thermal and Fluid Science **45** (2013) 203–212 21

[8] Stitou A. and Riethmuller M. L.: Extension of PIV to super resolution using PTV. Meas. Sci. Technol. **12** (2001) 1398–1403 21, 29

[9] Cavagna A., I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, V. Zdravkovic: The starflag handbook on collective animal behaviour: part i, empirical methods. Anim. Behav. **76** (2008) 217–236 21

[10] Takehara K., T. Etoh: A study on particle identification in ptv particle mask correlation method. J. Visualiz. **11** (1998) 313–323 21

[11] Shindler, Luca, M.M., Cenedese, A.: Using optical flow equation for particle detection and velocity prediction in particle tracking. Applied Mathematics and Computation **17** (2012) 8684–8694 21, 22, 28, 30

[12] Baker S., Matthews I.: Lucas-Kanade 20 years on: A unifying framework: Part 2. (2003) 22, 95

[13] Shi J., C. Tomasi: Good features to track. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (1994) 593–600 22

[14] Bradley A., A. Kruger, E.A.M., Muste, M.: Flow measurement in streams using video imagery. Water Resources Research **38**(12) (2002) 9, 22, 24

[15] Lecordier B., Demare D., Vervisch L. M., Riveillon J. and Triniti M.: Estimation of the accuracy of piv treatments for turbulent flow studies by direct numerical simulation of multi-phase flow. Meas. Sci. Technol. **12** (2001) 1382–1391 22, 24

[16] Gui L and Wereley S T: A correlation-based continuous window-shift technique to reduce the peak-locking effect in digital piv image evaluation. Exp. Fluids **32** (2002) 506–517 22

[17] Hauet, A., Willert, M. and Kompenhans, J.: Experimental system for real-time discharge estimation using an image-based method. J. Hydrol. **13** (2008) 105–110 22

[18] Nogueira J., Lecuona A., Ruiz Rivas U., and Rodriguez P.A.: Analysis and alternatives in two-dimensional multigrid particle image velocimetry methods: application of a dedicated weighting function and symmetric direct correlation. Meas. Sci. Technol. **12** (2002) 963–974 22, 24

[19] Eckstein A. C., Charonko J. and Vlachos P.: Phase correlation processing for dpiv measurements. Exp. Fluids (2008) 485–500 22, 25

[20] Eckstein A C, Vlachos P Digital particle image velocimetry (DPIV) robust phase correlation, Meas. Sci. Technol., **20** 2009, 485–500 22, 25

[21] Keane, R., Adrian, R.: Theory of cross-correlation analysis of piv images. Applied Scientific Research **49** (1997) 191–215 22

[22] Raffel, M., Willert, M., and Kompenhans, J.: Particle Image Velocimetry, A Practical Guide. (2007) 22, 27

[23] Huang H T, Fiedler H E and Wang J J Limitation and improvement of PIV: part I . Limitation of conventional techniques due to deformation of particle image patterns. Exp. Fluids, 1993 24

[24] Jambunathan K., Ju X.Y., Dobbins B.N. and Ashforth-Frost An improved cross correlation technique for particle image velocimetry. Meas. Sci. Technol. Volume 6. (1995), 507–514 24

[25] Astarita T. Analysis of weighting windows for image deformation methods in PIV. Exp. Fluids , **43** (2007), 859–872 24

[26] Theunissen R Theoretical analysis of direct and phase-filtered cross-correlation response to a sinusoidal displacement for PIV image processing. Meas. Sci. Technol., **23**(2012) 25

[27] Prasad A.K., Adrian R.J. Stereoscopic particle image velocimetry applied to liquid flows. Exp. Fluids, **15** (1993), 49–60 26

[28] Lecerf A., Renou B., Allano D., Boukhalfa A. and Trinite M. Stereoscopic PIV: validation and application to an isotropic turbulent flow. Exp. Fluids, **26** (1999),107–115 26

[29] Liu Z.C., Adrian R.J., Meinhart C.D. and Lai W. Structure of a turbulent boundary layer using stereoscopic, largeformat video-PIV. In: Adrian RJ, Durao DFG, Durst F, Heitor MV, Maeda M, Whitelaw JH, Developments in laser techniques and fluid mechanics (1997). Springer, Berlin Heidelberg, pp 259– 273 26

[30] Westerweel J, van Oord J: Stereoscopic PIV measurements in a turbulent boundary layer. In: Stainslaus M, Kompenhans J, Westerweel J(eds), Particle image velocimetry: progress toward industrial application. Kluwer, Dordrecht, 2008 27

[31] Hill DF, Sharp KV, Adrian RJ: Stereoscopic particle image velocimetry measurements of the flow around a Rushton turbine.Exp.Fluids,**29**(2000),478 - 485 27

[32] Prasad, A. K. Stereoscopic particle image velocimetry. Exp. Fluids, **29** (2000), 103-116. 28

[33] Adrian R. J., Westerweel J. Particle Image Velocimetry. Cambridge University Press (2011). 27

[34] Heas P., E. Memin, D. Heitz, and P. Mininni. Bayesian selection of scaling laws for motion modelling in images. In: International Conference on Computer Vision (ICCV09), Kyoto, Japan, October 2009. 31, 32

[35] Krajsek K. and R. Mester. Bayesian model selection for optical flow estimation. In: DAGM-Symposium. (2007), 142–151 31

[36] Wernet M Symmetric phase only filtering: a new paradigm for DPIV data processing. Meas. Sci. Technol., 2005, 601 - 618 25

[37] Richard P. Wildes, Michael J. Amabile, Ann-Marie Lanzillotto, and Tzong-Shyng Leu Recovering Estimates of Fluid Flow from Image Sequence Data Computer Vision and Image Understanding, **80** (2000), 246–266 35, 36

[38] Y. Nakajima, H. Inomata, H. Nogawa, Y.Sato, S. Tamura, K. Okazaki, and S. Torii Physics-based flow estimation of fluids Pattern Recognition, vol. 36, pp. 1203-1212, 2003. 35, 36

[39] B. Jahne and S. Wass, Optical wave measurement technique for small scale water surface waves, In: Proc. Advances in Optical Instruments for Remote Sensing, pp.147–152, 1989. 36

[40] H. Saikano, Fluid Motion Estimation Method based on Physical Properties of Waves, Comp. Vision and Pat. Recognition (CVPR), 2008 36, 37

[41] Zhengyou Zhang, Flexible camera calibration by viewing a plane from unknown orientations, In ICCV, 1999, 666–673 47, 55

[42] Hartley, R. I. and Zisserman, A., Multiple View Geometry in Computer Vision, Second Edition, 2004, Cambridge University Press, ISBN: 0521540518 xvi, 40, 42, 48, 59, 60, 62, 63, 70, 71

[43] Heikkila, Janne and Silven, Olli, A Four-step Camera Calibration Procedure with Implicit Image Correction, Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97), isbn = 0-8186-7822-4, 47, 51

[44] Levenberg, Kenneth (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares. Quarterly of Applied Mathematics. 2, 164 - 168. 50

[45] Trucco, Emanuele and Verri, Alessandro, Introductory Techniques for 3-D Computer Vision (1998), isbn = 0132611082, Prentice Hall PTR, Upper Saddle River, NJ, USA 55, 69

[46] A. Bradley ,A. Kruger, E. A. Meselhe and M. Muste, Flow measurement in streams using video imagery, Water Resourches Research, vol. 38, no. 12, 1315 , (2002) 6, 7, 8, 12

[47] R Tsubaki, I Fujita, S Tsutsumi, Measurement of the flood discharge of a small-sized river using an existing digital video recording system, Journal of Hydro-environment Research, 5 (2011) 313–321 6, 8, 10

[48] Muste, M., I. Fujita, and A. Hauet (2008), Large-scale particle image velocimetry for measurements in riverine environments, Water Resour. Res., 44, W00D19, doi:10.1029/2008WR006950. xv, 6, 7, 8, 9, 11, 12

[49] Kim T. K., J. K. Paik and B. S. Kang, Contrast Enhancement System using Spatially Adaptive Histogram Equalization with Temporal Filtering, IEEE Trans. on Consumer Electronics, vol. 44, 1:82-87, 1998. 7

[50] Wang M., J. Pan, S. Chen and H. Li, A Method of Removing the Uneven Illumination Phenomenon for Optical Remote Sensing Image, Proc. of IEEE International Geoscience and Remote Sensing Symposium, Korea, vol. 5, 3243-3246, 2005. 7

[51] Avgerinakis Konstantinos, Briassouli Alexia, Kompatsiaris Ioannis, Real time illumination invariant motion change detection, Proceedings of the First ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams, 2010, Firenze, Italy, p.p.75-80 7

[52] Fujita, I., and S. Komura, (1994), Application of video image analysis for measurements of river-surface flows (in Japanese), Proc. Hydraul. Eng.,38, 733-738. 7, 8

[53] Creutin, J. D., M. Muste, and Z. Li (2002), Traceless quantitative alternatives for measurements in natural streams, Hydraulic Measurements and Experimental Methods, Am. Soc. of Civ. Eng., Estes Park, Colo. 8, 10

[54] Fujita, I., Watanabe., H. and Tsubaki, R. (2007). Development of a non-intrusive and efficient flow monitoring technique: The space-time image velocimetry (STIV) Intl. J. River Basin Management Vol. 5, No. 2 (2007), pp. 105 -114 8, 10, 12

[55] Ichiro Fujita , Marian Muste and Anton Kruger (1998), Large-scale particle image velocimetry for flow analysis in hydraulic engineering applications, Journal of Hydraulic Research, 36:3, 397-414, DOI: 10.1080/00221689809498626 8, 9, 11, 12

[56] R. J. Adrian, Particle-imaging techniques for experimental fluid mechanics, Annu. Rev. Fluid Mech., 23, p.p. 261-304, 1991 9, 95

[57] Fujita, I. and Tsubaki, R. (2002). A Novel Free-Surface Velocity Measurement Method Using Spatiotemporal Images, Hydraulic Measurements and Experimental Methods, ASCE, on CDROM. 10

[58] J Nogueira and A Lecuona and P A Rodriguez, Data validation, false vectors correction and derived magnitudes calculation on PIV data, Measurement Science and Technology, 8(12), 1493, 1997 11

[59] I. Fujita and T. Kaizu, CORRECTION METHOD OF ERRONEOUS VECTORS IN PIV, Journal of Flow Visualization and Image Processing, 1995, 2, 173–185 11

[60] J.D. Creutin and M. Muste and A.A. Bradley and S.C. Kim and A. Kruger, River gauging using {PIV} techniques: a proof of concept experiment on the Iowa River, Journal of Hydrology, 277, 182 - 194, 2003 12

[61] M. Muste and H.-C. Ho and D. Kim, Considerations on direct stream flow measurements using video imagery: Outlook and research needs, Journal of Hydro-environment Research, 5 (4), 289 - 300, 2011 xv, 12, 13

[62] Jamo Ralli, PhD thesis, Fusion and regularisation of image information in variational correspondence methods, Universidad de Granada. Departamento de Arquitectura y Tecnologa de Computadores, 2011, http://hera.ugr.es/tesisugr/20702371.pdf 72

[63] Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool, SURF:Speeded Up Robust Features. Computer Vision and Image Understanding (CVIU).Vol. 110, No. 3, pp. 346 - 359, 2008. 88

[64] Baker Simon, Scharstein Daniel, J. P. Lewis, Stefan Roth, Michael J. Black, Richard Szeliski, A Database and Evaluation Methodology for Optical Flow, International Conference on Computer Vision - ICCV , pp. 1-8, 2007. DOI: 10.1109/ICCV.2007.4408903 91, 95

[65] Thielicke, W. and Stamhuis, E.J. (2014), PIVlab-Towards User-friendly, Affordable and Accurate Digital Particle Image Velocimetry in MATLAB. Journal of Open Research Software 2(1),30 95

[66] Thielicke, W. (2014), The Flapping Flight of Birds - Analysis and Application. Phd thesis, Rijksuniversiteit Groningen. 95

[67] Taylor, Z.J.; Gurka, R.; Kopp, G.A.; Liberzon, A., Long-Duration Time-Resolved PIV to Study Unsteady Aerodynamics, Instrumentation and Measurement, IEEE Transactions on , vol.59, no.12, pp.3262-3269, 2010, doi: 10.1109/TIM.2010.2047149 95