Université de Montréal

Cyclopean Optical Flow

par

Maria Fernanda Robles Hernandez

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Mémoire présenté à la Faculté des arts et des sciences
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en informatique

April 30, 2022

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

Cyclopean Optical Flow

présenté par

Maria Fernanda Robles Hernandez

a été évalué par un jury composé des personnes suivantes:

Liam Paull
(Professeur)

Max Mignotte
(Professeur)

Sébastien Roy
(Directeur)

Mémoire accepté le

RÉSUMÉ

Ce mémoire s'intéresse au domaine de la vision par ordinateur, et plus particulièrement l'estimation du flux optique. Le flux optique est un problème 2D notoirement difficile, car il est intrinsèquement sous-contraint. Pour introduire la notion de flux optique cyclopéen, nous allons considérer le problème en 1D pour éliminer le problème d'ouverture lié au mouvement 2D. Nous proposons une nouvelle approche basée sur un référentiel « cyclopéen », basée sur gradient calculé dans un espace continu pour résoudre le flux optique 1D. Ce mémoire se concentre a garantir que le gradient reste utilisable dans un intervalle suffisamment grand pour couvrir le déplacement spatial du mouvement. Lors de la résolution sur une approche coarse-to-fine, une représentation pyramidale est utilisée. Les résultats sur des images aériennes ainsi que des données synthétiques sont prometteurs. Ce travail se distingue des tendances actuelles en flux optique parle fait qu'il se spécialise pour les flux optiques à faible mouvement.

Nos résultats ont montré une bonne gestion des faux positifs tout en conservant une bonne densité. Nous considérons que la fiabilité des mesures de mouvement est très élevée, ce qui est au moins aussi important que la précision elle-même dans beaucoup d'applications. Ainsi, la polyvalence de la représentation "continue" permet de mieux contrôler la densité obtenue en fonction de la scène analysée. À notre avis, cette approche, qui complète les méthodes traditionnelles, ouvrira la voie à de nouvelles approches en apprentissage profond.

Mots clés:cyclopéen, vision stéréo, low ligne de base, pyramidal

ABSTRACT

This thesis is in the field of computer vision, focusing on the problems of optical flow estimation. Optical flow is a notoriously difficult 2D problem since it's inherently under-constrained. To introduce the concept of cyclopean optical flow, we will downgrade the 2D into 1D to make it more accessible. It proposes a new approach based on a "cyclopean" frame of reference. We apply a constrained gradient-based technique to solve 1D optical flow, for which the constraints are gradient behavior and correlation score. This thesis focuses on the fundamental problem of ensuring that the gradient remains usable in an interval large enough to cover the spatial displacement of motion. The proposed "cyclopean" approach does not enforce optical measurements over a fixed grid, which results in more reliable results. To further increase the allowed motion interval, we propose a pyramidal constraint that allows solving over a coarse-to-fine approach. We solved over aerial imagery, Sintel data-set, and Sintel data-set when artificially displaced 10% of the ground truth. This work is developed in the "continuous" framework commonly used for small motion optical flow. Our results showed good management of false positives while maintaining a good amount of convergence density. However, our method isn't as precise as the current state-of-the-art benchmarks, as it specializes in very small motions. Also, it's important to mention versatility comes with the concept of "continuous" representation. This allows us to select regions to be solved, opening the possibility of adapting to the spectrum of sparse or dense optical flow.

From this study point of view, we can highlight traditional methods have relevance even in the deep learning era, offering a new set of tools to exploit on the pursue of solving optical flow.

Keywords: Cyclopean, Stereo, Depth estimation, Small-baseline, pyramids

TABLE OF CONTENTS

LIST OF ACRONYMS & ABBREVIATIONS

---

nD	n Dimensional
CNN	Convolutional Neural Networks
SS	Small-baseline Stereo
LSM	Least Square Method
LK	Lukas-Kanade
H&S	Horn and Schunck
SAD	Sum of absolute differences
ZNCC	Zero mean normalized cross-correlation

ACKNOWLEDGMENTS

I'm extremely grateful for patience and lenience which make of the world a welcoming place; family, friends and acquaintances have greeted me in the context of this thought. I'm specially grateful to the university community: colleges, professors, administrators and my research director Sébastien Roy.

Chapter 1

INTRODUCTION

Optical flow is an important problem in the field of computer vision, therefore, research is highly motivated to address its challenges. Measuring apparent motion is the main goal of optical flow; In visual perception, the relative motion between an observer and a scene through time. In a digital context, optical flow is often a bridge between reality and computers; a camera is the source of observation, providing the necessary set of images.

1.1 Optical FLow

Optical flow is the computed motion given a sequence of images. To reach a solution, a final and initial state are defined by the nature of a video sequence V. As seen in Fig. 1.1, given two consecutive images I_1 and I_2 taken from V, the position of pixels is tracked and matched to similar distributions between I_1 and I_2.

The extraction of an image information will be directly linked to an algorithm's capacities, to which we can identify two extremes: local and global algorithms. The first category refers to pixel matching, a solution is calculated in a secluded manner; this technique highlight its robustness to noise. The second category involves all pixels to calculate displacement; these techniques provide a solution for each and every pixel of an image.

Adding complexity to optical flow algorithms can decrease deficiencies in the solution, however, an increase in the computation time is expected. This capacity adapts to the concept of sparse matching; this modality will only focus and compute solutions for

Figure 1.1: Optical flow estimation. A grid of instantaneous motion vectors is computed from a sequence of images. Left: I_1 and I_2, Right: I_1's optical flow.

certain pixels of interest. Alternatively, dense matching calculates every solution albeit and outcome with I_1's dimentions is computationally demanding.

To overcome optical flow's challenges in the most optimal way, the academic community has invested incredible efforts in defining one of the most dynamic frameworks.

Local linearization is used in the context of instantaneous motion, which reefers to small displacements between frames. Gradient-based techniques as categorized by Barron et al. commonly apply local linearization to match pixels between frames; the gradient's tangent behavior will be used to pursuit a pixel between frames, estimating their displacement.

To interpret displacement, information needs to be extracted from images to identify similarities and determine their variance in position. The information that will be explored in this chapter is the intensity of an image, which is the value of its pixels. Considering that the intensity is given as a discrete value, gradient-based estimation is of big interest to exploit the result of approximating it to a continuous function.

A gradient-based method requires two assumptions: two images have highly related distributions due to their sequential nature and the sequences of images have a constant and small interval of sampling that describe t and $t+\delta t$. These suppositions are known as *brightness constancy* which will give confidence to the analysis and supports that an image's intensity will conserve its brightness through changes in time, as a result, pixel tracking will be a constraint as follows:

$$I(\vec{x},\vec{y},t) = I(\vec{x}+\vec{u},\vec{y}+\vec{v},t+\Delta t) \tag{1.1}$$

For any pixel with an intensity value of $I(\vec{x}+\vec{u},\vec{y}+\vec{v})$ at $t+\Delta t$ exists as well in $I(\vec{x},\vec{y})$ at t.

This is considered to be a very rigid definition due to the equality symbol, which means no change in intensity is expected after a point of interest has been relatively moved in space. This doesn't consider the natural 3D behaviors of light reflection, perspective changes, or rotational movement, so the calculated displacement would be an approximation.

Gradient-based algorithms described and analised by Fleet and Weiss, where we can follow the mathematical process behind an 1D iterative gradient-based method. Intensity $I(\vec{x},t)$ is assumed to remain constant from one frame to another in spite of the motion,

$$I(\vec{x},t) = I(\vec{x}+\vec{u},t+1) \tag{1.2}$$

where $t = 1$ it's an ideal constant sampling rate. Any two pixels among the frames t and $t+1$ respecting this definition are presumed to be the same component in the scene with an apparent 1D motion of $\vec{u}$.

Considering two 1D intensity functions $I_1(x) = I(\vec{x},t)$ and $I_2(x) = I(\vec{x},t+1)$, the objective is to find u according to (1.2),

$$I_1(x-u) = I_2(x) \tag{1.3}$$

to gain access to $\vec{u}$, we use Taylor's series expansion, for which,

$$I_1(x-u) = I_1(x) - uI_1'(x) + O(u^2 I_1''(x)) \tag{1.4}$$

and therefore, we can write u (x's difference between signals),

$$I_1(x) - I_2(x) = uI_1'(x) + O(u^2 I_1''(x)) \tag{1.5}$$

where an approximation of u is defined,

$$u \approx \frac{I_1(x) - I_2(x)}{I_1'(x)} \tag{1.6}$$

by avoiding the high degree derivatives from Taylor's expansion. This is then improved with an iterative approach in the form of Gauss-Newton optimization,

$$u^n \simeq u^{n-1} + \frac{I_1(x+u^{n-1}) - I_2(x)}{I_1'(x+u^{n-1})} \tag{1.7}$$

where n is the number of iterations. In the formulation (1.7) each step of n we explore the function $I_1(x)$ until 1.2 is achieved. By the risk of falling into an infinite loop of iteration, n is usually limited by a maximum amount of iterations allowed.

But as mentioned, Horn and Schunck's method is an approximation limited by its own initial hypothesis, although its results are sustainable, it welcomes development in order to increase performance and robustness, as followed by [12, 9].

1.1.1 Horn and Schunk

The most fundamental constraint of optical flow estimation is the *constant brightness assumption*. It stipulates that the image pixel intensities are preserved during motion, so we can say

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$

for a pixel undergoing motion $(\Delta x, \Delta y)$ in time interval Δt. Assuming the motion and the sampling time are small, we have the relation:

$$
\begin{aligned}
&I(x,y,t)-I(x+\Delta x,y+\Delta y,t+\Delta t)=0\\
&\frac{\partial I}{\partial x}\Delta x+\frac{\partial I}{\partial y}\Delta y+\frac{\partial I}{\partial t}\Delta t=0\\
&(\frac{\partial I}{\partial x},\frac{\partial I}{\partial y})\cdot(\frac{\Delta x}{\Delta t},\frac{\Delta y}{\Delta t})=-\frac{\partial I}{\partial t}\\
&\nabla I\cdot\vec{v}=-I_t
\end{aligned}
$$

This *constant brightness* equation can't fully resolve the 2D motion $\vec{v}$.

Therefore, Horn and Schunck's (H&S) variational method introduces *brightness consistency* and global smoothness regularization into optical flow as seen in the cost function Eq. 1.8, which enables pixel-to-pixel comparison while maintaining a system with a linear behavior.

$$[h]E=\iint_{x,y}(f_x u+f_y v+f_t)^2+\lambda(u_x^2+u_y^2+v_x^2+v_y^2)dxdy \tag{1.8}$$

The smoothness constraint assumes that pixels' motion will spatially change slowly, the second term of Eq. 1.8 represents the motion vectors $\{u,v\}$, which are expected to be minimized. Also, the term λ will offer weight to either assumption, favoring the smoothness regularization if large.

By using differential calculus, we isolate the unknowns $\{u,v\}$ from Eq. 1.8,

$$
\begin{aligned}
u&=u_{avg}-f_x\frac{P}{D}\\
v&=v_{avg}-f_y\frac{P}{D}
\end{aligned}
\tag{1.9}
$$

where,

$$
\begin{aligned}
P&=f_x u_{avg}+f_y v_{avg}+f_t\\
D&=f_y^2+f_x^2+\lambda
\end{aligned}
$$

Eq. 1.10 are applied iteratively to estimate the motion vectors $\{u, v\}$,

$$\begin{aligned} u^n &= u_{avg}^{n-1} - f_x \frac{P}{D} \\ v^n &= v_{avg}^{n-1} - f_y \frac{P}{D} \end{aligned} \tag{1.10}$$

where,

$$\begin{aligned} P &= f_x u_{avg}^{n-1} + f_y v_{avg}^{n-1} + f_t \\ D &= f_y^2 + f_x^2 + \lambda \end{aligned}$$

This method is also constrained by a 2D projection as the representation of a 3D physical world, making estimation more accessible without highly compromising accuracy during the solution's approximation. This algorithm is categorized as gradient-based because it uses the Spatio-temporal derivatives of a scene's sequence.

1.1.2 Lucas and Kanade

Lucas et al. (LK) solved the aperture problem by assuming it's probable that an object's displacement will be alike among its pixels, allowing to solve 1.11, which is now overconstrained. With this new assumption, we introduce n number of equations; n represents the neighborhood which we assume is within the object.

$$C = \sum_n (f_{xi} u + f_{yi} v + f_{ti})^2 \tag{1.11}$$

We solve then the system of n equations for $\{u, v\}$,

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_n f_{xi}^2 & \sum_n f_{yi} f_{xi} \\ \sum_n f_{xi} f_{yi} & \sum_n f_{yi}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_n f_{ti} f_{xi} \\ -\sum_n f_{ti} f_{yi} \end{bmatrix} \tag{1.12}$$

with least squares method (LSM).

Again, this method is bounded by the previous assumptions, where we benefit of small motions.

1.1.3 Coarse-to-fine approaches

Pixels can be hard to densely match when displacements are not uniform through the images. An alternative to this issue is coarse-to-fine (pyramidal representation) [2] introduces some context to the calculations by analyzing different resolutions of an image with the intention of filtering high frequencies. This has a low-pass filter behavior, allowing the algorithm to deal with images without many details, being obstruction when trying to match two distanced pixels. This categorizes methods as global optimization and has become a recurrent aid for a variety of methods [21, 54, 55, 49].

1.1.4 Neural networks

Under the inspiration of the previous work, different methods have been proposed over the years [12, 9, 39] expanding optical flow's traditional background; Neural Networks' application to optical flow exhibits a new wave of discoveries aiming to make self-driving cars, futuristic medical tools, surprising virtual and augmented reality systems, etc, a reality. Deep learning offers context extraction and cleaner results, controlling the current state of the art [59, 56, 31, 51].

1.1.5 Challenges of Optical Flow

Computing optical is not always easy, or possible at all. Common assumptions are: intensity does not change with motion (Lambertian objects, static scene), image is assumed locally continuous, small sampling time, etc. This results in a major constraint: motion must be small. However, since the matching is not discrete but continuous, sub-pixel motions are easy to compute.

LK and H&S are methods restrained by the above, therefore their pure application is very limited to small motion. Coarse-to-fine is a good complementary technique to increase other methods' capacity but, it's limited too. In certain circumstances, there is a loss of information due to smoothing over small details and the possibility of false convergence if the shape of the curves does not translate adequately through the pyramidal levels.

For the development of an optical flow algorithm, it's important to study the complications behind previous work, to mention a few:

- Brightness inconsistency: An object with different illumination conditions tend to result in different intensities. Lambertian or "matte" surface will have apparent brightness consistency due to the lack of specular reflection.

- Occlusions: When elements of a scene disappear between frames due to movement, it's referred to as an occlusion. Optical flow encounters challenges with objects moving out of the frame of sight or having a three-dimensional (3D) rotation.

- Ill-textured surfaces: A repetitive pattern increases the appearance ratio of the same intensity pixels; this error is classified as false-positive. Lack of texture also is a difficult zone due to the little amount of information available.

- Ground-truth data: in order to develop optical flow methods, ground-truth data is essential to analyze an algorithm's performance. Currently, the available data sets are scarce, and extracting data from sensors, cameras, etc. can be an expensive task being focused on autonomous driving. Alternately, simulated data sets also provide quality scenes with a diversity of movements for benchmarking.

Variational methods find themselves limited by the *brightness consistency*, and the simplification of this approach reduces the capacity for the algorithms to catch up with Deep Learning's achievements. This method also struggles with big displacements, Gradient-based pixel-to-pixel matching is not enough to constraint/capacitate the method to converge into fake solutions.

The coarse-to-fine methodology might highly change shape during scale reduction. This leads to convergence in false positives due to error propagation. Also, if an image is too detailed, pyramid representation might filter this information away and therefore lose information for those instances.

Deep learning methods are especially disadvantaged by the lack of variety and quantity of data during the training process. The sensitivity during their learning process is also a vulnerability because the risk of over-fitting is hard to control due to the lack of labelled data.

In our method, we are going to expect these limitations, so we intend to solve small motions, and explore how these can become larger in a context where the image is assumed to behave smoothly.

We do not intend to build robustness to image noise, bias, specularity, or other mismatches of intensities. However, our method will not use the temporal derivative, so there is no need to assume that the image flow, in time, is continuous. Therefore, it's possible to do stereo since images in a stereo pair are "discrete" and feature depth discontinuities would be incompatible with a continuous temporal image sequence.

1.1.6 From One Dimensional Optical Flow to Stereo

In Chapter 2 we will introduce and apply cyclopean optical flow's concepts to solve small-baseline stereo (SS) matching for aerial imagery.

When a pair of images have proper epipolar geometry, optical flow from I_1 to I_2 will be parallel to one of the axes. With this definition, we can apply 1D optical flow to solve displacements between I_1 and I_2, as seen in [15, 58, 13].

Stereo matching requires a pair of images taken simultaneously (stereo pair), but in the case of aerial imagery, *motion stereo* is the available data (consecutive images from a translating camera). On that account, the *motion stereo* pair will be translated as a SS for the cyclopean stereo matching algorithm.

1.2 Small baseline stereo

The main characteristic of stereo is the use of two cameras displayed horizontally. In an account of the latter, the two pairs of images are projected in the same geometric plane (co-planar) and have a geometric disparity between them. In case images are not co-planar, they can be rectified into a shared plane. Some methods to re-project such images are: planar, cylindrical, and polar rectification with epipolar geometry.

Some algorithms [13, 15, 58] have instigated optical flow techniques for depth estimation. A big number of stereo algorithms assume camera calibration and epipolar geometry.

In [14, 26] it's said that a big angle between views is generally expected to reach good precision, in contrast, [14] SS is proposed. The latter assumes that disparity between two images varies slowly in space and reinforces that small angles (which generate fewer occlusions and higher similarity in a stereo-pair) favor the estimates of depth.

During SS 2 cases are identified:

- Small displacements due to SS.
- Big displacements due to far away elements.

The two cases are recurrent in aerial imagery because elements are really distant from the viewer and *motion stereo*; the latter translate into SS, where the baseline is extremely small due to *motion stereo*, where stereo-pairs are actually consecutive images of a video taken from a translating camera, for example, Fig. 1.2.

In stereo, displacement of objects is essentially 1D optical flow along epipolar line and the displacement is directly related to depth. Here we explain this application and how we solve it with the 1D cyclopean flow.

Optical Flow is the perceptible two-dimensional (2D) motion of a scene through time. Having a pair of images, we track similarities in order to estimate displacement.

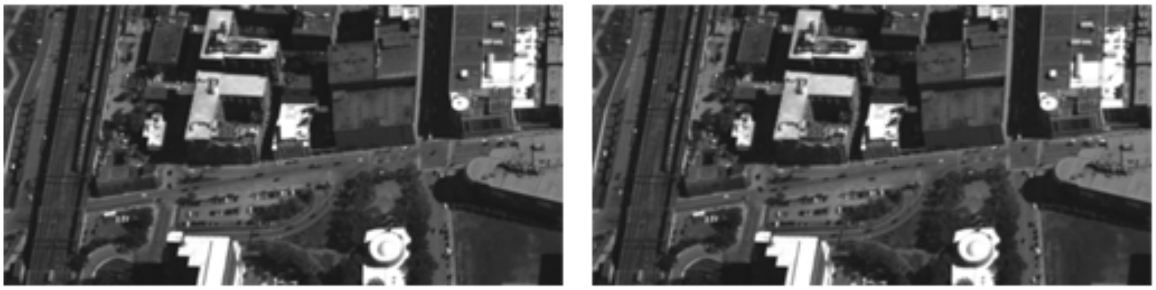


Figure 1.2: Motion stereo pair with epipolar geometry with size of 800 × 400 pixels

This problem is approached by a variety of techniques with the goal of overcoming recurrent difficulties during its calculation, for example: obtaining real-world ground-truth data, scenes susceptible to brightness, occlusions, etc.

According to Hur and Roth, Shah and Xuezhi, the current state of the art is defined by Convolutional Neural Networks (CNN).

The potential of neural networks was defined by its early performers [59, 18, 3, 23, 24, 48] stating the capacity of deep architectures to resolve optical flow. Baseline work can be described by two main trains of thought: CNN feature extractors and end-to-end CNN regressors.

An image's extracted features enhance the representation to be employed by correspondence techniques resulting in a more informed process to achieve pixel tracking, these models [59, 23, 24] became a confident alternative to traditional concepts by describing a pixel not only by its intensity but also by its attributes, tackling natural brightness inconsistencies.

Meanwhile, the future of nonimage pre-processing wouldn't be ensured until further research was made on end-to-end CNN regression architectures [18, 48] and reveal a competitive performance in [32, 55].

Introducing optical flow as a purely learning problem became a viable option for motion estimation. And so forth, developments in the subject keep opening interest to refine the initial surge of neural network methods.

Despite the latter progress, deep learning encounters limitations that also motivate further research in the traditional domain of optical flow, some examples reside in inconvenient training time, data generalization, and lack of data sets for training. Non-learning algorithms still show relevance due to their robustness, independence from training, and potential collaboration with other motion estimation techniques to enhance a solution as seen in [3, 23, 24].

Earlier work still motivates current advances in the field becoming a strong background for optical flow. Such is the case for the method proposed in this paper, which follows gradient-based techniques as categorized by Barron et al.

Gradient-based methods make part of the main foundations for practical applications in optical flow [21, 19, 51], where an image's intensity and it's spatio-temporal derivatives are processed for pixel tracking. These methods rely on the *brightness consistency* assumption, where perceived elements of a scene will display constant intensity across frames; despite the ideal implementation of optical flow, accuracy is still achieved. Nevertheless, reducing the complexity of light behavior urges some weaknesses, being: as occlusions and large displacements.

The lack of representation capacity introduces a vulnerability commonly known as the aperture problem. Horn and Schunck (H&S) proposes a global smoothness term, where regularization will take into account all pixel's flow computation to be minimized. This approach motivated further research, being improved or used as a foundation for new methods.

Lucas et al. (LK) approaches the aperture problem in a different manner, the regularization is local, considering only a small neighborhood N to constraint the equations of optical flow. The least-square method (LSM) is applied to minimize motion for each pixel in N, finding an optimized displacement solution. This method is also the object of inspiration for advances in the field.

Correlation-based methods maximize similarities between data. [38] introduces these algorithms as Spatio-temporal users, therefore related to gradient-based techniques. Sun highlights the application of these techniques is fast, including: the sum of absolute differences (SAD), a simple and fast approach, and zero-mean normalized cross-correlation (ZNCC), the best candidate for measuring similarities between areas. in spite of optical flow and stereo similarities, the first is not greatly influenced by correlation measures.

Nevertheless, hand-made algorithms have limitations and it's essential the usage of image processing tools to enhance accuracy. Reducing the scale of an image is a recurrent concept used in computer vision to alter information with the goal of obtaining a preferable version of the image's signal [2]. A pyramid is a multilevel image representation that computes n number of sub-samplings from a reference I. The motivation behind this decision is that a simpler form of I will be more stable due to its smoother shape. Also, the construction of multiple images at various scale levels combines the information of a larger spatial neighborhood. A set of scaled versions of I is known as pyramidal representation and its application is categorized as fine-to-coarse practice as seen in [49, 48, 55, 8].

In order to introduce cyclopean optical flow, our immediate goal isn't to solve 2D displacements, We plan to lower the dimension proportion to implement over 1D displacements. This is a gradient-based method that measures correspondences between two images, and the direct implementation of 1D cyclopean flow targets small-baseline stereo (SS) correspondence for depth estimation, which we compare to 1D motion.

A stereo-pair image is a set of two images representing the same scene from two slightly different points of view. This strategy is bio-influenced by human vision, in which we can perceive the similarities of double image capture. Depth estimation benefits from stereo arrangements because it's possible to talk about depth estimation in terms of "correspondence". SS targets the stereo captures for which the two points of view are separated by a very small distance.

Initial interest in Stereo vision has provided countless materials for further development in this field. The context of this article was mainly built around a traditional framework [50, 57, 7, 28] emphasizing the traditional approach which our method is associated to.

Cyclopean depth estimation assumes *brightness consistency* in a local pixel matching process. Other optical flow techniques have been implemented through optimization [13, 15, 58] showing that minimization is viable during stereo matching. Our matching method is an iterative gradient-based which follows the convergence behavior of a Gauss-Newton optimization.

Similar to other areas of computer vision, the efforts in deep learning have conquered public benchmarks. CNN has shown relevance in this field [11, 25, 62, 45], but unlike its optical flow application, results still struggle to accurately correlate points during stereo matching. Research move towards different network designs, architectures oriented to cost aggregation [63, 60] which offers a competitive computational cost while still using features, end-to-end models [36, 64].

But as mentioned in [63], traditional elements are still used, therefore, the potential remains unexplored in hand-made methods.

1.2.1 Small stereo Benchmarks

Benchmarks such as MPI Sintel [10], KITTI-2015 Optical Flow [43], Middlebury [4] are fundamental for the development of optical flow, as mentioned before, access to data-sets is pretty limited. Their integration and availability brand them as references for evaluating the functionality of algorithms in challenging scenarios, for example: long sequences, large motions, and highly physics-ruled scenes. The Sintel data-set extracts ground-truth motion from synthetic scenes, which in spite of being simulated, respects the distribution of real-world situations captured on clips. Sintel feature large motions, so the images are ajusted to feature the small motions required by our method. The KITTI data-set is a compilation of driving films with extensive tracking as navigation or data gathered by a 3D laser scanner to obtain the ground truth. The Middlebury dataset is a collection of different real-world scenes with ground-truth measurements.

Chapter 2

ONE DIMENSIONAL CYCLOPEAN FLOW

This chapter presents our new 1D optical flow method, Cyclopean optical flow, and its use for solving Small-baseline Stereo.

This research is focused on improving the estimation of optical flow. We propose to use as "cyclopean value", which is an intermediate reference point between two signals. The cyclopean value is expected to increase accuracy due to its nature of sharing a distribution with all signals involved in the process of solving optical flow.

With this approach, we intend to increase the capacity of gradient-based methods by having two dynamic ends to describe motion. Additionally, the algorithm will also be capable to be constrained through the same dynamism.

Also, the use of pyramidal representation will be explored to enhance results with a smooth transition of cyclopean optical flow through low and high resolution. Coarse-to-fine will be approached as a generator of initial values and as a constraint function, therefore, we don’t plan to introduce the weaknesses mentioned above.

2.1 Image representation

An image is a common visual representation given by pixels and their distribution but this representation might become deficient

when computing optical flow due to derivative computation; accordingly, derivatives are notoriously sensitive to noise and can be inaccurate when computed over discrete values, being images a natural source of these defects. To tackle this problem, image interpolation is proposed to transform pixels into a representation that features continuity in its values and first derivative.

The interpolation form applied is a cubic spline interpolation where $I(x) = \ddot{I}(x)$ to respect cardinality constraints. From now on, I will refer to the interpolation of any image $\ddot{I}$. The degree of this spline grants enough capacity to ensure continuity on the first and second derivatives. To avoid instability at the function's edges, padding is recommended to ensure the influence of the original representation still precedes. This addition does not affects the method greatly, it just avoids loss of information.

It's possible to see in Fig 2.1 that interpolation of first-order yields discontinuous first derivatives. Discontinuity results in a low-resolution derivative, which can translate into rough changes along with spatial exploration which won't allow an accurate perception of the image's shape. Also, the low capacity to represent the gradient of an image will have a higher probability to land in problematic situations as described in subsection 2.2.3.

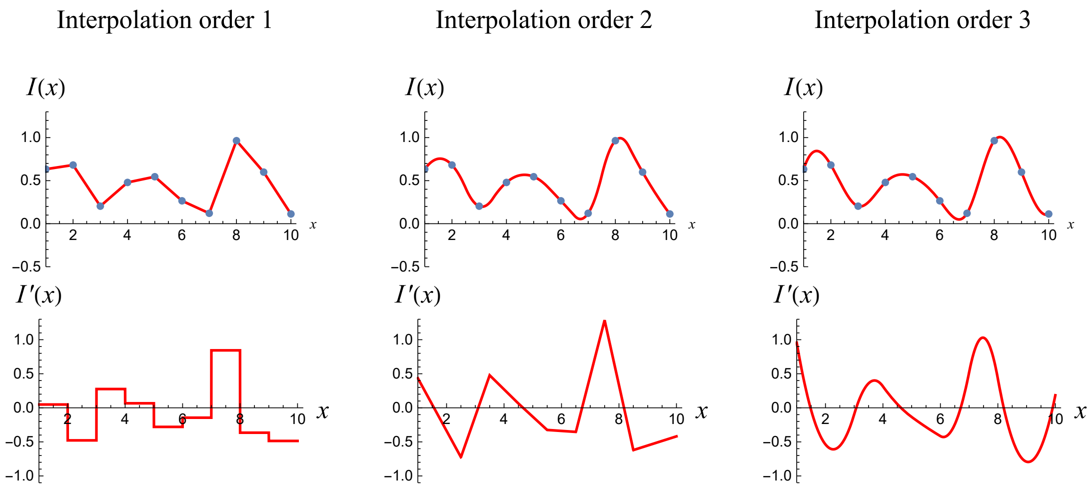


Figure 2.1: Cubic spline interpolation. Top row is image intensity, bottom is first derivative. Most Left: Linear interpolation (1st order). Most right: Cubic interpolation (3rd order).

Note that the cyclopean method only relies on spatial interpolation, not temporal interpolation. Time is considered discrete, allowing single dependency on spatial derivatives.

In everything that follows, it will be assumed that the image used is in fact a continuous interpolation of the original discrete image.

2.2 Cyclopean Optical Flow

1D Cyclopean optical flow is a technique to estimate 1D displacement in an image's intensity when its scene is affected by motion. In a sequence of 2 images, the first element (I_1) defines the pixel's original state, and the following segment (I_2) will acquire I_1's altered distribution due to the existing motion dynamics. The theoretical average intensity that would be observed in the middle of the transition from I_1 towards I_2 is defined as a "cyclopean image" (C). Using C's intensity values as a reference allows decomposing the optical flow solution into two parts: the flow from I_1 to C and the flow from C to I_2. Both velocities are then added up and are assumed to represent the movement behavior between I_1 and I_2.

Considering a continuous representation of C, I_1 and I_2 (further details in section 2.1), the cyclopean image is computed as follows:

$$C(x) = \frac{I_1(x) + I_2(x)}{2} \tag{2.1}$$

The accuracy of this model is expected to increase due to the inclusion of every image derivative instead of a single one. Including both ($I_1'(x)$ and $I_2'(x)$) the method traces more information making it possible to identify unusual behavior and avoid finding fake solutions. In section 2.2.3 this issue will be addressed in greater depth. This model is also enhanced due to the computation of two smaller motions instead of a large one. The introduction of two displacements artificially increases the sample rate, allowing to capture more details on the scene's transition offering a smoother solution. Figure 2.2 illustrates the traditional and cyclopean methods and the properties mentioned are graphically identifiable.

This is also mathematically perceived through their optical flow definitions. Traditional optical flow is computed as:

$$v = \frac{I_1(x) - I_2(x)}{I_1'(x)} \tag{2.2}$$

The equation 2.2 is derived from a Taylor series approximation, it only considers the spatial derivative of I_1 to estimate the displacement v. On the other hand, cyclopean optical flow is computed as:

$$\begin{aligned} v &= v_1 + v_2 \\ v_1 &= \frac{I_1(p) - C(p)}{I_1'(p)} \\ v_2 &= \frac{C(p) - I_2(p)}{I_2'(p)} \end{aligned} \tag{2.3}$$

Having two estimations influencing v's value constrains the result offering a more reliable result.

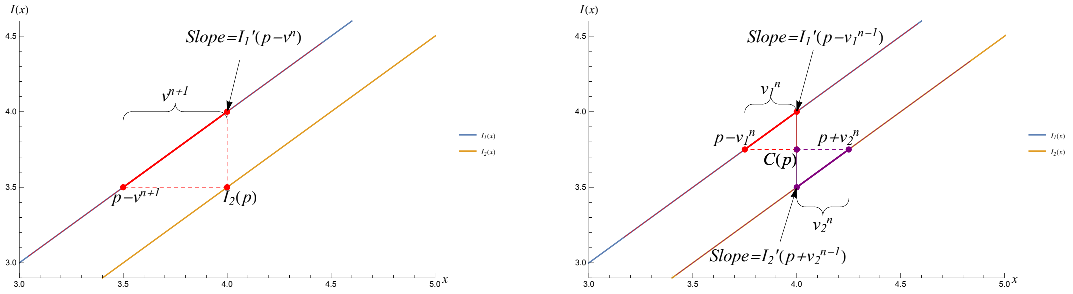


Figure 2.2: Optical flow between images I_1 and I_2. Left: traditional optical flow. Right: Cyclopean optical flow.

It's important to mention that the cyclopean solution of point p is equivalent to the solution of $p - v_1$, therefore, all motion vectors will be represented with an initial start point of $p - v_1$.

2.2.1 Iterative estimation

According to the traditional method, optical flow can be computed exactly when the starting point and the destination point have the same gradient, as seen on 2.2. The spatial derivative acts as the slope of a straight line which will be constrained by $p - v^{n+1}$ and $p - v^n$ in x and due to the euclidean behavior of the system, the right distance v is found immediately.

In the cyclopean method, the gradient of I_1 must be the same between point p and $p - v_1$, and for I_2 it must be the same at p and $p + v_2$; nevertheless, this gradient behavior is the attribute of a basic grayscale and it does not appear often in practice. Non-linear cases are more common systems where the gradient changes spatially and the solution would be just an approximate, making it more challenging to find a good solution.

To increase accuracy in non-ideal cases, it's proposed to estimate v with an iterative approach. Starting from an initial guess, the optical flow will be updated along with its gradient until convergence.

Equation (2.4) is used to update and approximate v_1 through iteration, with v_1^n representing v_1 at step n, and an initial value of $v_1^0 = 0$. This represents the search of c_p's intensity in $I_1(x)$. The process will become redundant once $p - v_1^n$ satisfies $I_1(p - v_1^n) = c_p$, where p's theoretical pixel has been successfully tracked and found in $I_1(x)$.

$$v_1^n = v_1^{n-1} + \frac{I_1(p - v_1^{n-1}) - C(p)}{I_1'(p - v_1^{n-1})} \tag{2.4}$$

The same computation is performed for v_2, with the update function (2.5).

$$v_2^n = v_2^{n-1} + \frac{I_2(p + v_2^{n-1}) - C(p)}{-I_2'(p + v_2^{n-1})} \tag{2.5}$$

The first and second iteration steps are shown in fig. 2.3. It's observed that the initial gradient is located at point p, and v_1^1 isn't the exact solution due to the spatial changes in the gradient. The second iteration updates the values of v_1^1 using the gradient at point $p - v_1^1$ which results in a much more accurate motion.

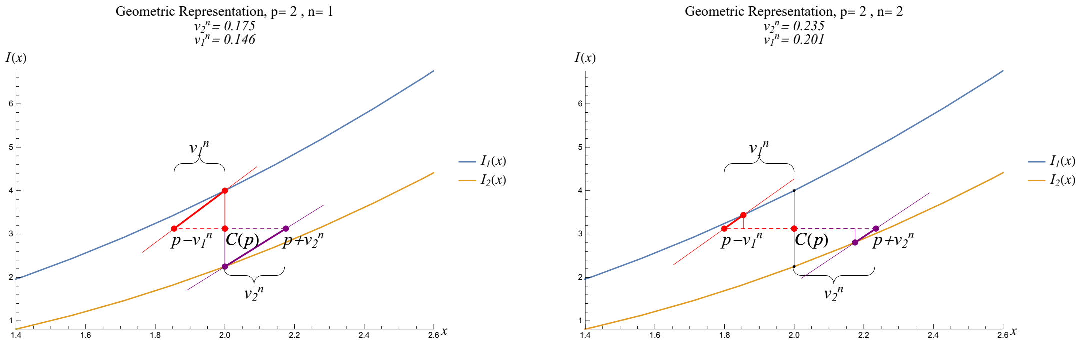


Figure 2.3: Cyclopean optical flow to solve v_2 and $v2$. Left: $n = 1$, right: $n = 2$

2.2.2 *Initial values*

This section presents the selection of an initial value for the iteration. We also show how stochastic selection can be used to enhance convergence.

The goal of an initial value is to approach the method performance towards the concepts that ensure an admissible solution.

Due to the nature of cyclopean method, an initial value v^0 will be composed by two elements: v_1^0 and v_2^0. The capacity of v^0 is precised by a suitable starting range $v^0 \in \mathbb{R} : b \leq v^0 \leq a$ translated to an equivalent proposition of v_1^0 and v_2^0 where $v^0 = v_1^0 + v_2^0$ and $v_1^0 \times v_2^0 \geq 0$. During the translation, $v_1^0 \in \mathbb{R} : b \leq v_1^0 \leq a$ and $v_2^0 \in \mathbb{R} : b \leq v_2^0 \leq a$. Fig. 2.4 represents the euclidean visualization of v^0.

The values of v^1 and v^2 share sign because otherwise, it's interpreted as if the c_p is not located in the intermediary space $I_1(p)$ and $I_2(p)$, which contradicts the definition of c_p.

This approach haves an objective to provide a roughly estimated range of initial values yet enough capacity to describe cyclopean distance.

We define the values of v_0 with the expected minimum and maximum flow solution.

Figure 2.5 introduces convergence in terms of initial values v_1^0 and v_2^0, where

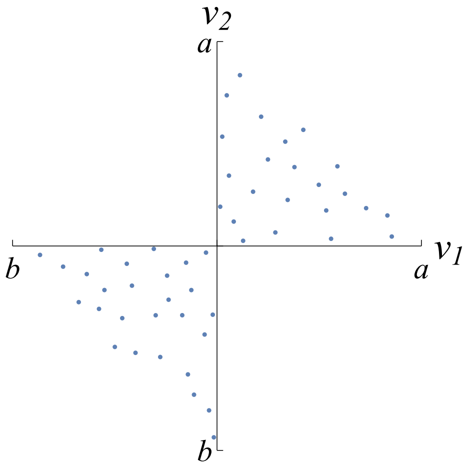


Figure 2.4: Initial values in range of $[a, b]$ described by v_1^0 and v_2^0.

$v_1^0 \in \mathbb{R} : -2 \leq v_1^0 \leq 2$ and $v_2^0 \in \mathbb{R} : -2 \leq v_2^0 \leq 2$ for $p = 11$. The solution displayed also satisfy gradient and correlation constraints, of which, conditions are raised in subsections 2.2.3 and 2.2.4. Well estimated solutions will appear along the ground truth line $v_1^0 + v_2^0 = 1$ and will converge if started in the convergence zone. The latter is the effective combination of v_1^0 and v_2^0 to converge into $v = 1$, revealing the ideal initial values range as: $v_1^0 \in \mathbb{R} : -0.5 \leq v_1^0 \leq 1.5$ and $v_2^0 \in \mathbb{R} : -0.5 \leq v_2^0 \leq 1.5$; any initial value with an ideal definition should converge into an accepted solution, as seen in Fig. 2.6.

2.2.3 Gradient Constraints

As presented above, using an iterative approach can solve optical flow when the gradient is changing spatially.

However, there are more cases to consider in order to compute optical flow robustly.

The following section will discuss some convergence issues related to gradient behavior and how we propose to tackle them.

Figure 2.7 illustrates the ideal cases. When a gradient is fixed over the motion, a sin-

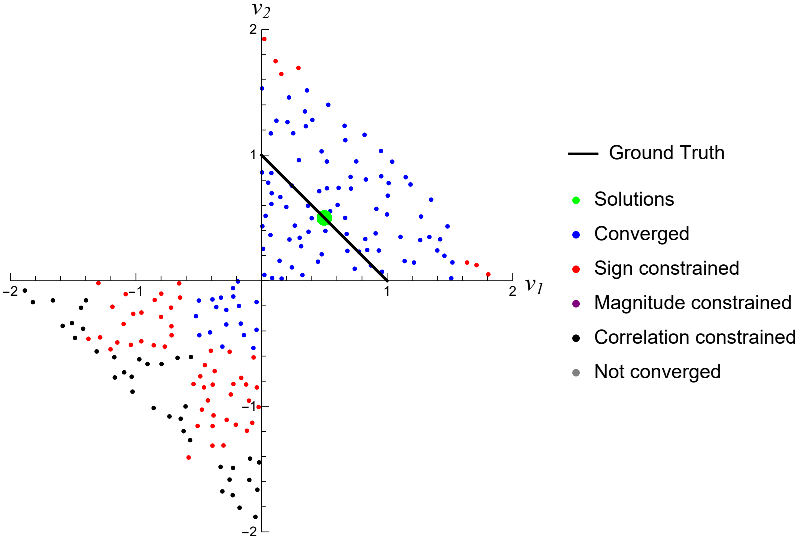


Figure 2.5: Convergence in terms of v_1^0 and v_2^0.

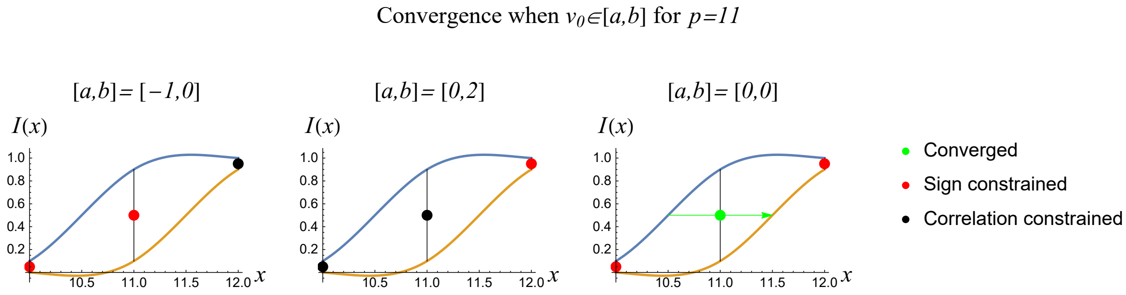


Figure 2.6: Right: convergence zone solution. Middle and Left: not in convergence zone solutions.

gle iteration provides the exact flow (Fig. 2.2). When the gradient is changing spatially, iterations are needed but also result in a correct solution (Fig. 2.3). A big characteristic of this situation is that I_2 duplicates I_1 in the range of displacement without breaking any 1D physical grounds, this method might encounter some limitations at inflection points which from a short-sighted perspective would seem I_1 and I_2 are not the same images.

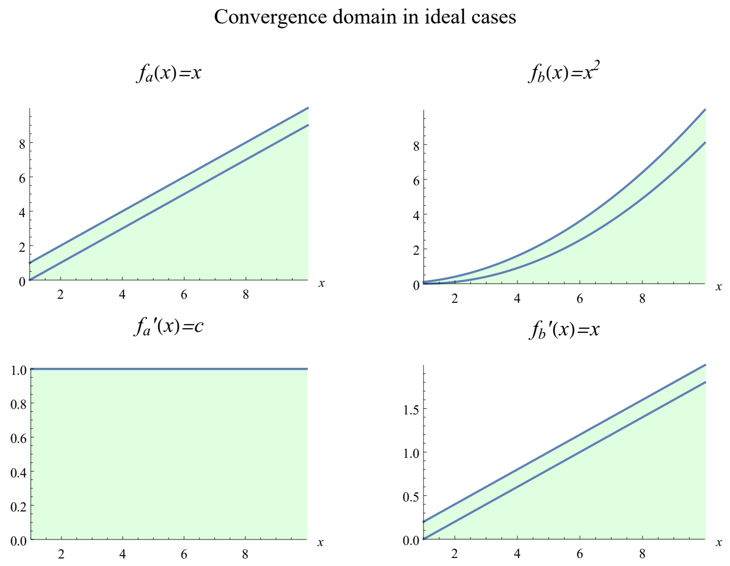


Figure 2.7: Ideal gradient behavior. Left: fixed value solved with single iteration. Right: varying gradient solved with multiple iterations

Sign of gradient

It's reasonable to expect that at convergence $I_1'(p-v_1) = I_2'(p+v_2)$, meaning that I_1 has a cyclopean displacement of $v = v_1 + v_2$ at point p in an ideal way. Being the gradients equal at p, they share signs and it's assumed that gradients will keep the same sign throughout the iterations. If it happens otherwise, it's considered that the optical flow cannot be estimated at p, since it probably won't converge.

Case 1 ("Same sign") is illustrated as Fig. 2.8 and then defined as:

$$v^n = \begin{cases} v^n = v_1^n + v_2^n, & \text{if } I_1'(p - v_1^n) * I_2'(p + v_2^n) > 0 \\ undefined, & \text{otherwise} \end{cases}$$

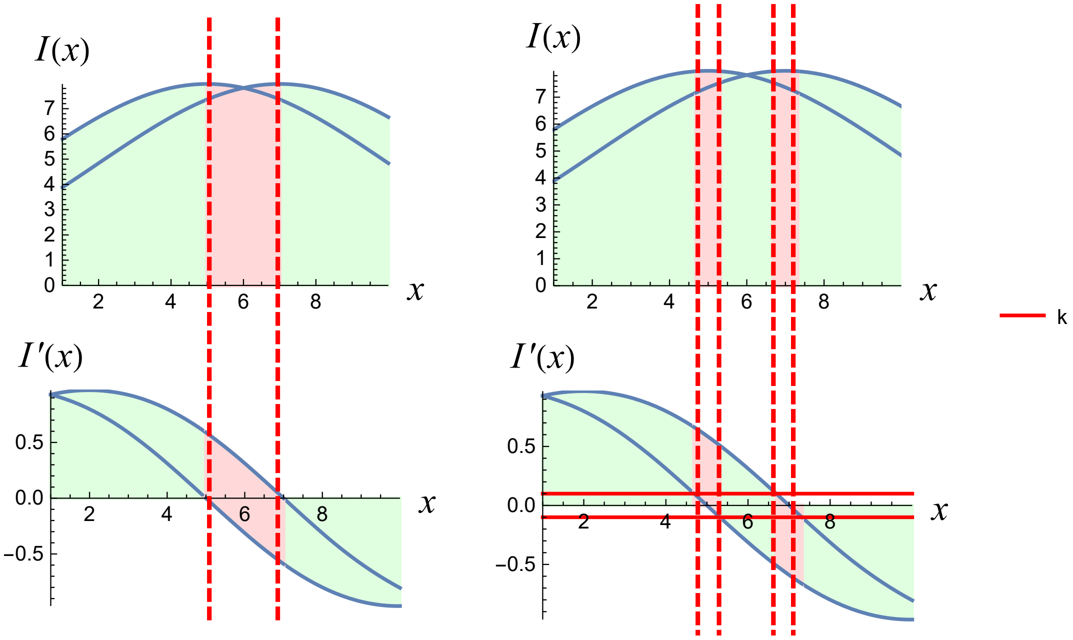


Figure 2.8: Case 1. Green interval: ideal behavior. Red interval: no possible solution due to sign mismatch. Case 2. Green interval: ideal behavior. Red interval: no possible solution due to a small gradient magnitude ($k = 0.05$).

Magnitude of gradient

The magnitude of the gradient is a denominator in equation 4.4, therefore, a small gradient would result in a big estimate of motion impacting negatively on the computation. The low gradient is to be considered as a lack of texture in images which isn't enough information to compute optical flow. A threshold k will be introduced to define the lowest value of gradient the method can accept in order to have a good approximation.

This behavior will be considered as Case 2 ("Magnitude of gradient"), illustrated as Fig. 2.8 and defined as:

$$v^n = \begin{cases} v^n = v_1^n + v_2^n, & \text{if } |I'(p)| > |k| \\ undefined, & \text{otherwise} \end{cases}$$

2.2.4 Correlation Constraint

Pearson's correlation is applicable during stereo matching, because the application fits in the test's linearity assumption. In Fig. 2.9 we have a global representation of linearity where p is displaced to p_1 and p_2. If motion is properly estimated, $I_1(p_1)$ should be linearly related to $I_2(p_2)$. Each point p varies in displacement.

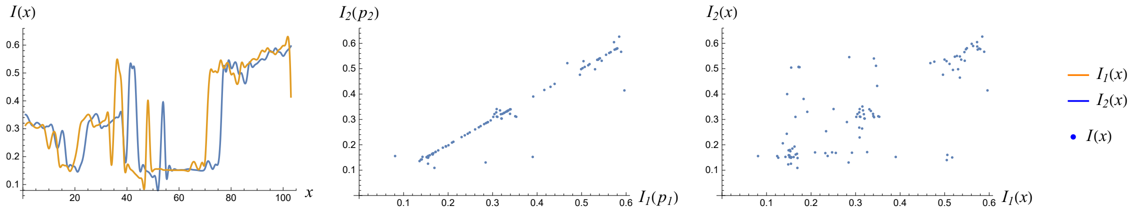


Figure 2.9: $I_1(x)$ vs $I_2(x)$. Middle: variables linearity when corresponding displacement p. Right: variables non linearity when no displacement p.

But this property is only applicable in a local manner, where p's displacement is assumed to be uniform in order to evaluate linearity.

To implement this constraint we use two intensity vectors z described as:

$$z_1(n, \Delta\, n) = \{I_1(p_1 - n), ..., I_1(p_1 - \Delta n), I_1(p_1), I_1(p_1 + \Delta n), ..., I_1(p_1 + n)\} \tag{2.6}$$

$$z_2(n, \Delta\, n) = \{I_2(p_2 - n), ..., I_2(p_2 - \Delta n), I_2(p_2), I_2(p_2 + \Delta n), ..., I_2(p_2 + n)\} \tag{2.7}$$

where n is in the range of a small neighborhood and Δn are constant steps towards n. Then (2.6) and (2.7) are evaluated on Pearson's correlation coefficient r, as consulted in Freedman et al.

The coefficient haves a score range of $-1 \geq r \geq 1$ where 1 is an ideal linear correlation, any score $0.5 \geq r \geq 1$ is considered to be a large linear correlation.

As seen in Fig. 2.9, we expect a positive linear behavior, translated in terms of Pearson's correlation as $0.5 \leq r$, therefore, during any iteration, we find:

$$v = \begin{cases} v = p_2 - p_1, & \text{if } r(z_1, z_2) \geq 0.9 \\ undefined, & \text{otherwise} \end{cases}$$

Note that this is just an acceptance criterion that will decide if a computed motion is coherent to its neighbors. In comparison to optical flow's algorithm LK, the constraint does not enforce a common motion to neighbor pixels, therefore, depth discontinuities will be preserved.

Images with too much information should have a smaller n, because the details might have non shared motion, therefore the assumption for correlation constraint would fail.

2.2.5 Enforcing constraints

Enforcing constraints haves as a goal:

- "No false positive": If the method converges, the answer is real.
- "No false negative": This is not possible in practice. There is a constraint on the image intensity, so the gradient can be unsuitable for the iteration.

If we enforce all previous constraints, it's clear that some points will remain unsolved, but we expect precision for those that are solved. The method will operate in the spectrum of dense to sparse solutions depending on how tight the constraints are. In order to estimate optical flow for all points, we will then need to "fill the gaps" by using other methods, for example, pyramidal representation.

2.2.6 Post-processing

If using a list of initial values (as seen in subsection 2.2.2), a list of solutions is expected. Therefore, we implement selection criteria to discard false positive results, as seen in equation 2.8. The latter it's composed of comparison operators, as the restrictions are supposed to reject excessive unusual behavior; boundaries are roughly defined as to filter values between a lower (b) and upper (a) range . Also, when applied in stereo, the direction can be restricted by only allowing results in the same coordinate space. We usually use the same minimum and maximum as the initial values definition.

$$C = \{v_1, v_2 |\ v_1 * v_2 > 0 \ \& \ a \geq v_1 + v_2 \geq b\} \tag{2.8}$$

2.3 Experiments

We will work on a synthetic signal as seen in fig. 2.10, which represents a interpolated step displaced $\Delta x = 1$ to validate the cyclopean method.

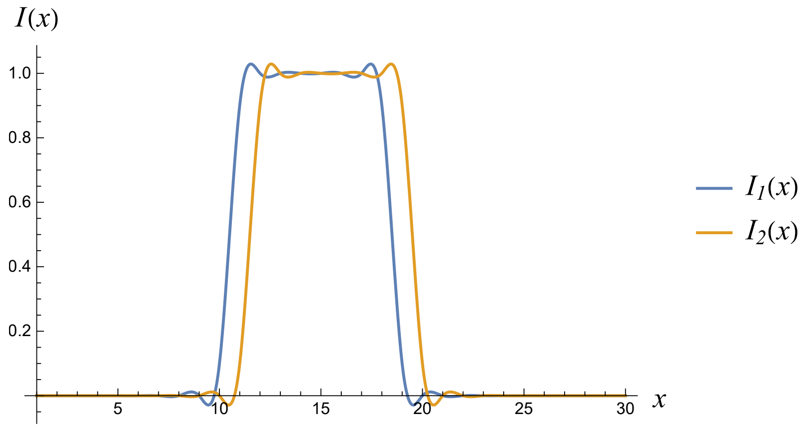


Figure 2.10: Interpolated step signal.

For the experiments we consider:

$$\begin{aligned} & v^0 \in \mathbb{R} : -3 \leq v^0 \leq 3 \\ & k = 0.001 \\ & z(n, \Delta\, n) = z(2, 0.2) \\ & C = \{v_1, v_2 |\ v_1 * v_2 > 0\ \&\ 0 \geq v_1 + v_2 \geq 2\} \end{aligned} \tag{2.9}$$

Figure 2.11 represents the convergence map of $p = 11$; The convergence zone is not only defined as: $v_1^0 \in \mathbb{R} : 0 \leq v_1^0 \leq 0.5$ and $v_2^0 \in \mathbb{R} : 0.5 \leq v_2^0 \leq 1$ but also as: $v_1^0 \in \mathbb{R} : 0.5 \leq v_1^0 \leq 1.5$ and $v_2^0 \in \mathbb{R} : 0 \leq v_2^0 \leq 0.5$ manifesting two solutions for $p = 13$.

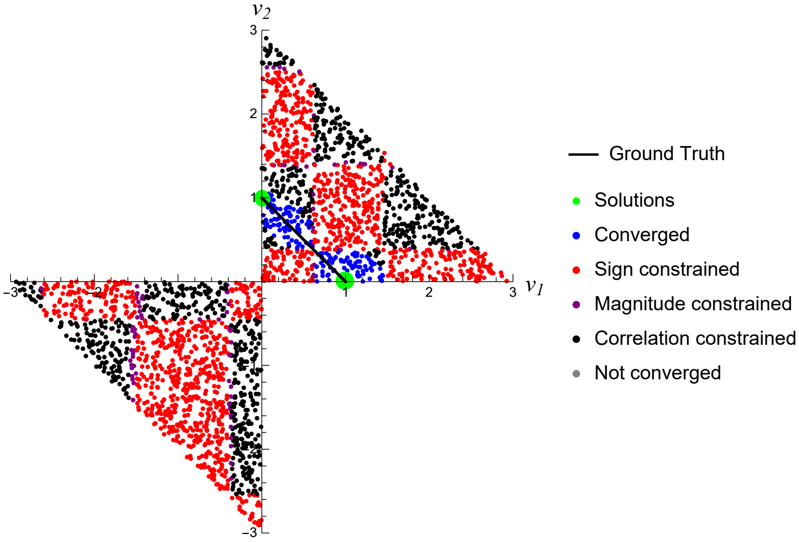


Figure 2.11: Convergence in terms of v_1^0 and v_2^0 of $p = 13$.

This behavior happens due to the interception of signals $I_1(x)$ and $I_2(x)$ in $p = 11$, therefore the cyclopean value C_p will overlap with the initial values of $I_1(p)$ and $I_2(p)$ having two initial candidates as a reference. The solutions will converge depending on the initialization values as seen in Fig. 2.12, where $v_n = 0$ means $I_n(p)$ was considered as a reference value.

Technically $[a, b] = [0, 0]$ could also be a solution, but in this case, the constraints of sign and correlation reject this convergence.

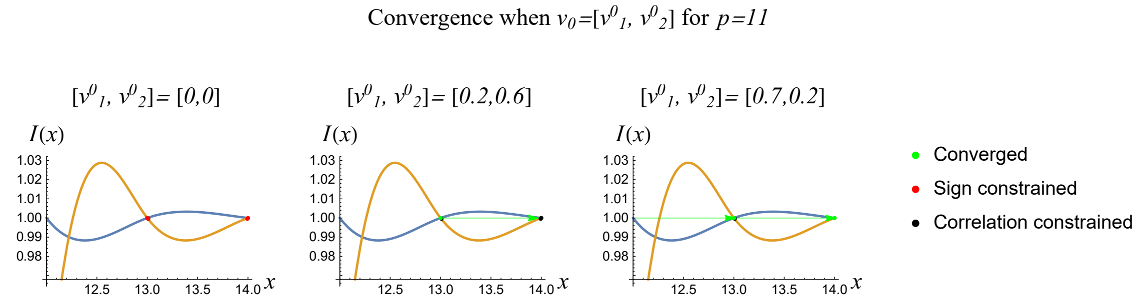


Figure 2.12: Right: constrained solution. Middle and Left: convergence zone solutions.

This case also depends on having an ideal initialization values, due to the reduced size of convergence zone.

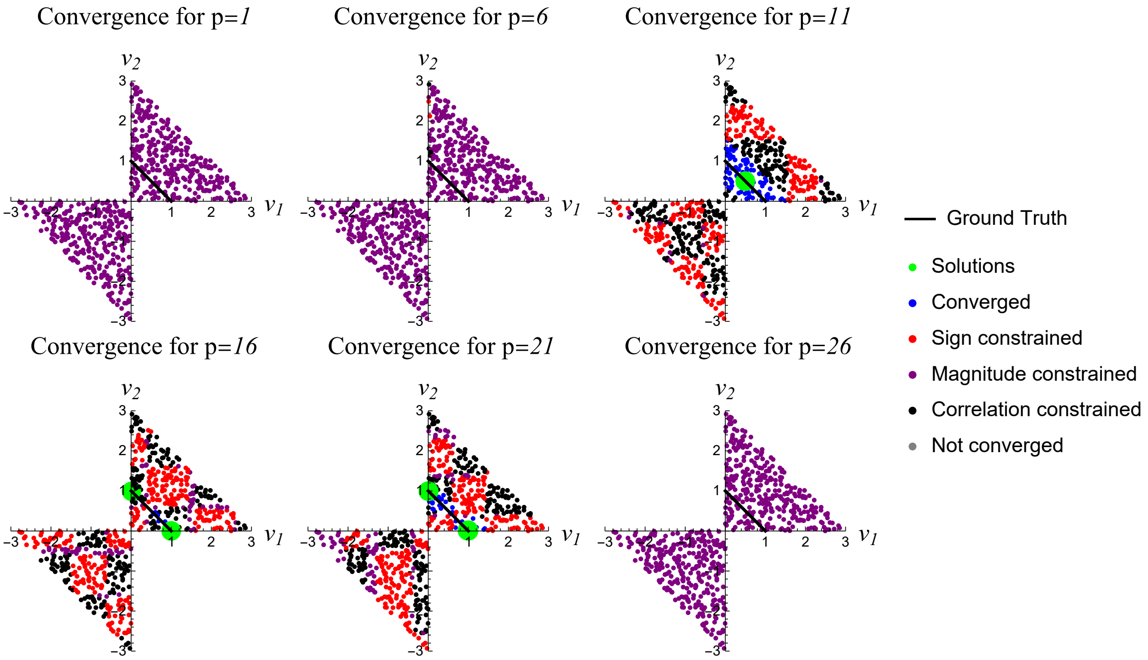


Figure 2.13: Convergence in terms of v_1^0 and v_2^0 for multiple points.

Further solutions in Fig. 2.13 ($p = 16,\ 21$) show a similar behavior convergence behavior. Meanwhile, the edges of the signal (ex. $p = 1, 6, 26$) will repeatedly be constrained by the "Magnitude constraint" (subsection 2.2.3). The algorithm is designed to distrust texture-less sections, even if ideally initiated, as seen in 2.14, therefore this points have no solution.

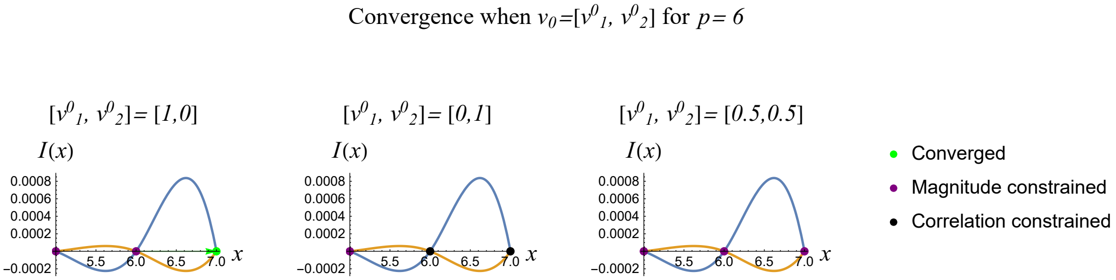


Figure 2.14: Magnitude constraint in $p = 6$.

The solution in $p = 11$ haves an ideal-like composition, where the convergence zone is wider and unique. In figure 2.15 we present two constrained solutions, initialized inadequately, for the algorithm to notice "sign" and "correlation" constraints.

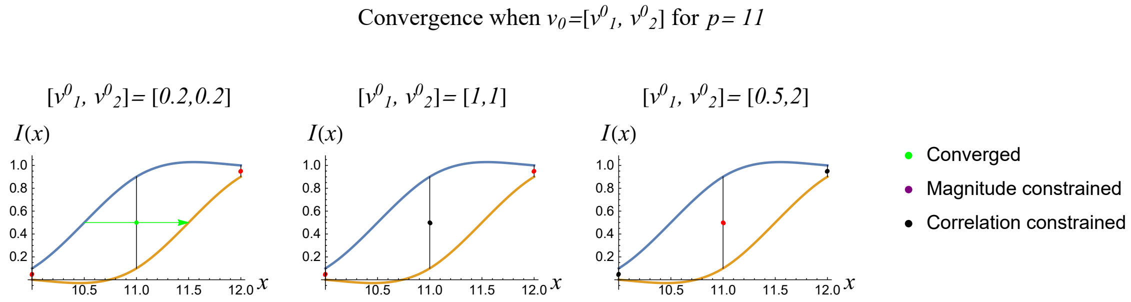


Figure 2.15: Right: solution in convergence zone. Middle and left: constrained solutions.

We add random noise of ±0.05 as seen in 2.16 and obtain the solutions for the points in Fig. 2.17. Same points are solved without noise as seen in Fig. 2.18.

The edges ($p = 1, 4, 7, 22, 25, 28$)have much more of correlation issues whenever noise is present because there's no information other than the noise, which is random.

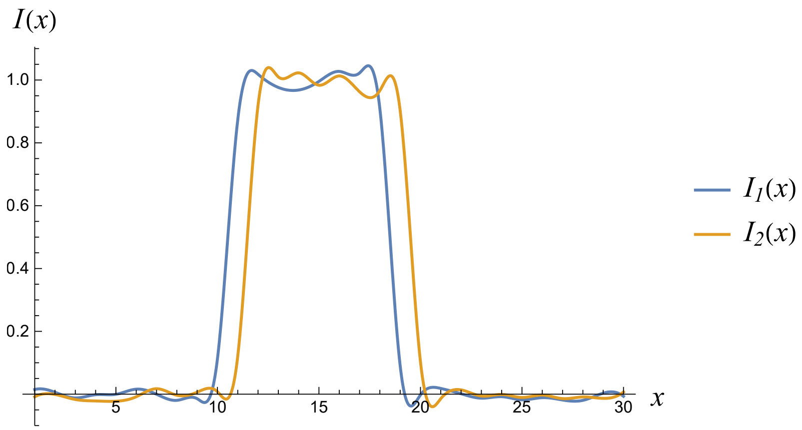


Figure 2.16: Interpolated step signal with simulated noise.

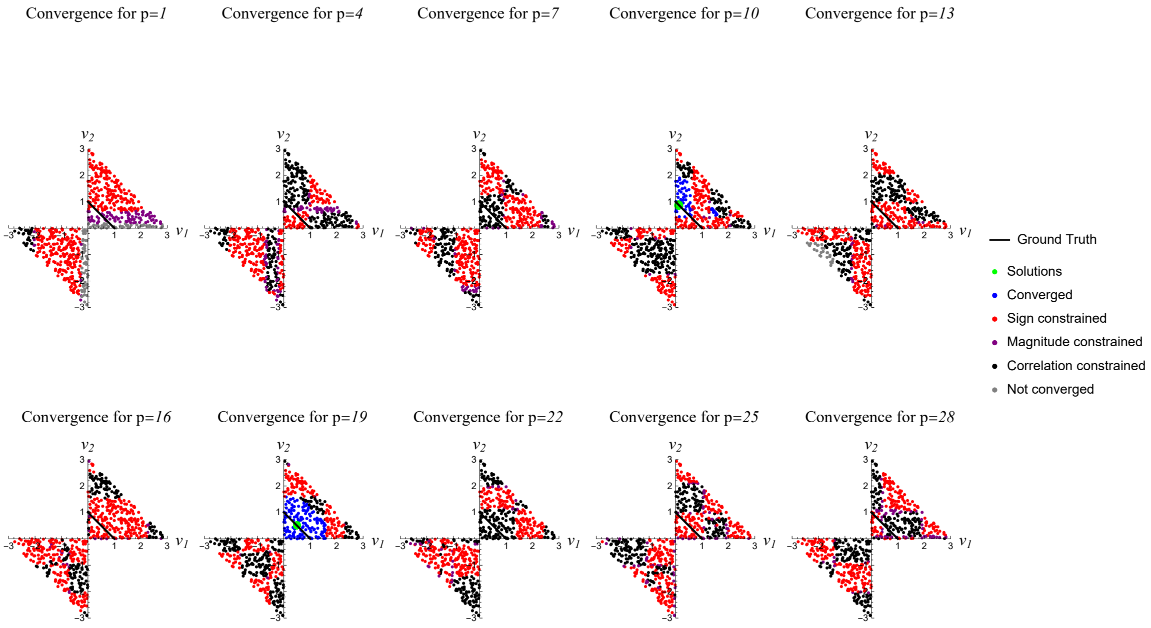


Figure 2.17: Convergence map of signal with simulated noise.

Through this examples it's highlighted that noise harms the method, but still manages to converge wherever the signal shape is closer to ideal ($p = 10, 19$) as said in subsection 2.2.3.

In the example of Fig.2.19, we can identify the method struggles to find a solution

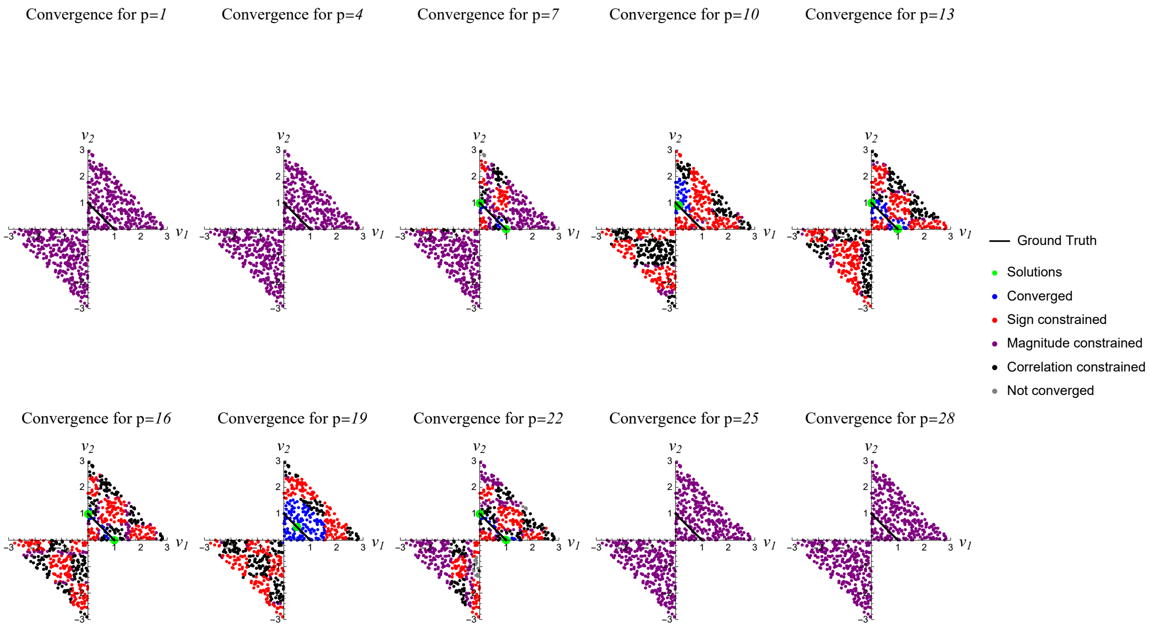


Figure 2.18: Convergence map of signal.

whenever noise is present in texture-less regions.

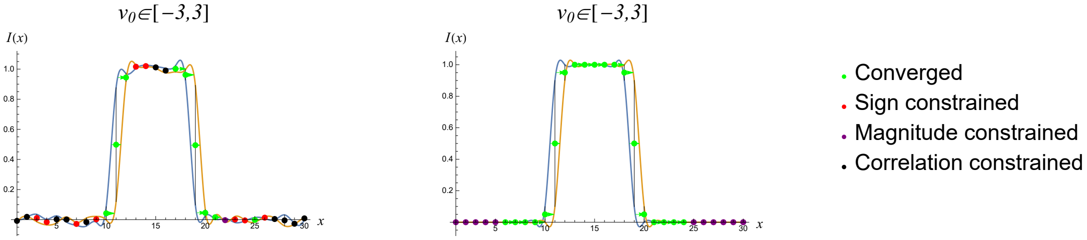


Figure 2.19: Convergence of signal, Left: Signal with noise, Right: Pure signal.

Chapter 3

CYCLOPEAN STEREO FOR SMALL-BASELINE STEREO

This chapter presents the application of our 1D cyclopean optical flow algorithm to the problem of 1D small-basline stereo.

Stereo matching proceeds along an single line, provided by epipolar geometry

3.1 Finding Epipolar geometry

In order to perform stereo matching the epipolar geometry must be known. Usually, stereo algorithms assume that images are rectified according to the estimated epipolar geometry in order to make epipolar lines horizontal, so the matching can proceed horizontally.

3.1.1 Aerial Stereo imagery

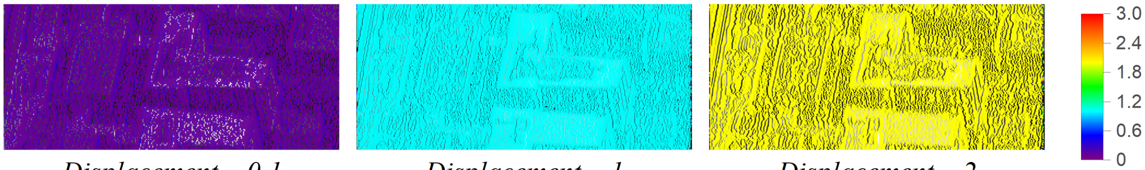
In the specific case of aerial imagery, one can assume that the camera is far from the scene, essentially providing a "small baseline" relative to the scene. This results in parallel epipolar lines, so all that is needed to rectify the stereo image pair is to estimate the rotation of the lines and perform a rectifying rotation.

In a stereo pair, the perceived motion is along epipolar line, so we assume that measuring motion along the wrong epipolar geometry will provide larger displacements. We will thus perform a 1D interval minimization to find the rotation that will provide the motion field with global minimal magnitude. In the Fig. 3.9, the minimum motion field was obtained for a rotation of 1.519 degrees of the second image. Figures 3.10 and 3.11 illustrate why this rectification is crucial to obtain got results.

3.2 Solving stereo

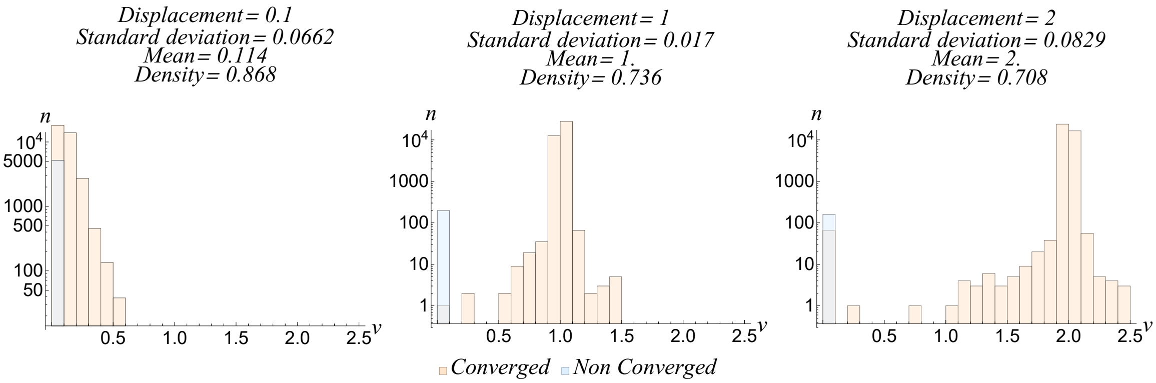
Assuming a rectified stereo pair, we expect all motions from one image to the next to be horizontal, and their magnitude is related to the depth of the objects. Measuring 1D optical flow is ideal for small baselines or when objects are far from the camera. Because of depth discontinuities, we consider our method to be well suited for this flow estimation since it does not require temporal continuity.

We start experimentation by examining the algorithm's performance over a pair of images $I(x)$ and $I(x+v)$, where $I(x)$ haves a synthetic displacement of v, as seen in Fig. 3.1. Initial values have a range of $v_0 = [0,3]$ and the selection criteria is defined as $C = \{v_1, v_2 | v_1 * v_2 > 0\&0 \geq v_1 + v_2 \geq 3\}$, $k =$ 0.001 and correlation vectors $z(2, 0.2)$.



Displacement = 0.1 *Displacement = 1* *Displacement = 2*

Figure 3.1: Motion stereo pair solved for different displacements



Displacement = 0.1
Standard deviation = 0.0662
Mean = 0.114
Density = 0.868

Displacement = 1
Standard deviation = 0.017
Mean = 1.
Density = 0.736

Displacement = 2
Standard deviation = 0.0829
Mean = 2.
Density = 0.708

Figure 3.2: Histogram of motion stereo pair solved for different displacements

During pixel tracking, our method uses the image's gradient to determine which position is best to explore and enhance the probabilities of $I(p \pm \triangle p) = C_p$. We are especially expecting good results on small displacements due to the model's high dependence on a signal's shape. Big displacements will result in a sub-optimal range of C_p values and the vulnerability of landing in zones where the real solution is not reachable due to non-desirable gradient behavior. Accordingly, Fig. 3.2 describes the results of optical flow over multiple displacements; the smallest displacement 0.1 presents the best performance in density and average error (width of the histogram). During these experiments, restrictions (Selection criteria, initial value) are ill-defined.

When providing a list restrained by different ranges, accuracy is expected to be higher when properly bounded. During these experiments, selection criteria are ill-defined. So, with a synthetic displacement of 0.1 we obtain the values in Fig. 3.4. This result introduces a decrease in standard deviation when the values are closer to the solution $v = 0.1$. Average and density exhibit a weak improvement through the different range values, indicating robustness regarding the initial state. Is to be concluded that values closer to the solution will favor the performance, but, in real case scenarios, this will rarely happen. Therefore, for further experimentation, we will use an initial range of $v_0 = [0, 1]$, in which we propose direction and a wide range of initial values. For further visualization, Fig. 3.3.

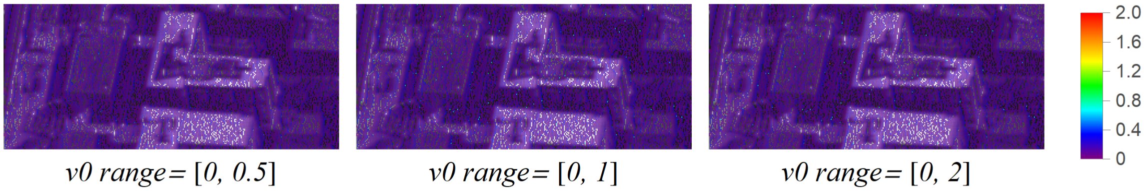
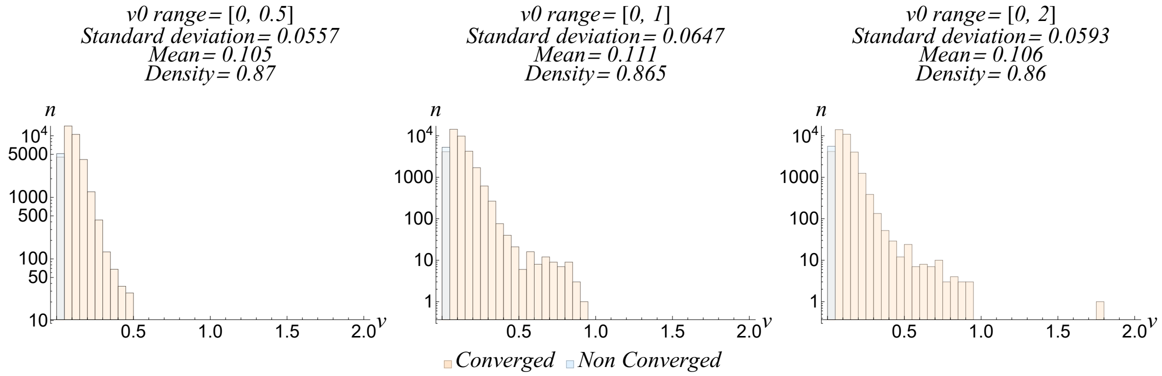


Figure 3.3: Motion stereo pair solved for displacements different initial value ranges

The selection criteria is a factor that will affect the results too. In the histograms Fig. 3.4 it's possible to see the method is fairly robust and the results don't relatively vary. If we look at the table, even if the difference is small, it can be identified that



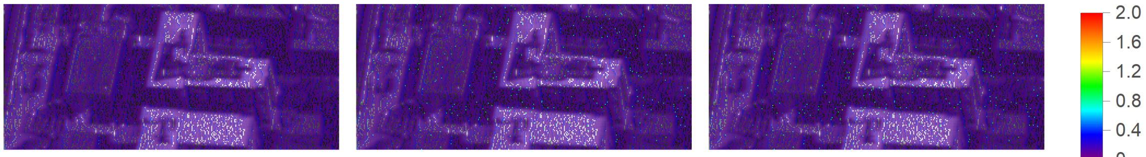
v0 range= [0, 0.5]
Standard deviation= 0.0557
Mean= 0.105
Density= 0.87

v0 range= [0, 1]
Standard deviation= 0.0647
Mean= 0.111
Density= 0.865

v0 range= [0, 2]
Standard deviation= 0.0593
Mean= 0.106
Density= 0.86

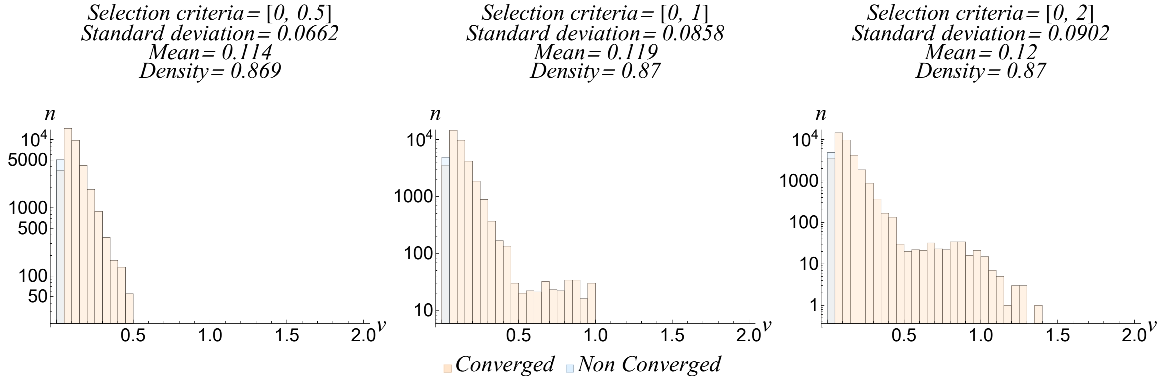
Figure 3.4: Histogram of motion stereo pair solved for initial value ranges

the criteria closest to the final solution will present less error, standard deviation with a very small penalty in density.



Selection criteria=[0, 0.5] *Selection criteria*=[0, 1] *Selection criteria*=[0, 2]

Figure 3.5: Motion stereo pair solved for different criteria



Selection criteria= [0, 0.5]
Standard deviation= 0.0662
Mean= 0.114
Density= 0.869

Selection criteria= [0, 1]
Standard deviation= 0.0858
Mean= 0.119
Density= 0.87

Selection criteria= [0, 2]
Standard deviation= 0.0902
Mean= 0.12
Density= 0.87

Figure 3.6: Histogram of motion stereo pair solved for different criteria

The synthetical experiments show a theoretical amount of robustness when in an ideal scenario. The factor with the most theoretical impact is the selection criteria subsection 2.2.6, which rejects ill-defined out-layers.

3.3 Experiments

3.3.1 Test 0: Aerial imagery

When we solve optical flow for Fig. 1.2 we obtain Fig. 3.7. Initial values have a range of $v_0 = [0,2]$, the selection criteria is defined as $C = \{v_1, v_2 | v_1 * v_2 > 0,\ 0 \geq v_1 + v_2 \geq 2\}$, $k =$ 0.001 and correlation vectors $z(2, 0.2)$. The ground truth is unknown, but based on context we conclude the following points:

- The top of the buildings will tend to have a disparity higher than 0.6
- Any section other than a ceiling gains magnitude under 0.6

This respects the physical property of perspective, where far away objects move faster than closer ones. Also, the lack of texture is affecting the convergence density, with only 27.1% pixels of the image converged.

Figure 3.7: Cyclopean optical flow solution of aerial imagery

Next we introduce the available aerial stereo pair 3.9 with of dimensions $1280x720$. The pair is rectified to align the epipolar geometry by a rotation of 1.519 degrees on

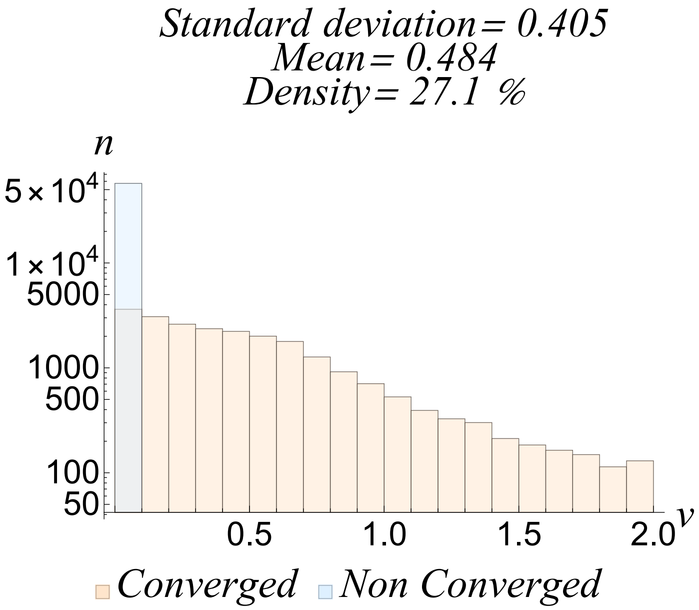


Figure 3.8: Cyclopean optical flow solution histogram of aerial imagery

the second image. We have then a fully solved image, where we enforce the discussed constraints, having as a result a partially sparse disparity map Fig. 3.10.

The same tests are attempted in a non rectified pair resulting in Fig. 3.11, where the solution is not coherent presenting a big variation between a surface's depths values, although it should be more uniform.

3.3.2 Sintel Dataset

Test 1: Sintel data-set $I_1(x)$ is artificially displaced $I_1(x - d * 0.1)$ where d is the ground truth; maximum displacement is $v_{max} = 3.86$. Initial values have a range of $v_0 = [0, -1]$, the magnitude constraint is $k = 0.001$, the correlation vector is defined as $z(5,\ 0.2)$ and the selection criteria is defined as $C = \{v_1, v_2 | v_1 * v_2 > 0,\ 0 \geq v_1 + v_2 \geq -1\}$. The upper bounds are set to only focus on small movements. We obtain Fig. 3.12 and Fig. 3.13.

Test 2: Sintel data-set $I_2(x)$ is artificially displaced $I_2(x + d * 0.9)$ where d is the

Figure 3.9: Aerial stereo pair. Notice that the second image has been rectified to align its epipolar geometry.

ground truth; maximum displacement is $v_{max} = 3.86$. Initial values have a range of $v_0 = [0, -1]$, the magnitude constraint is $k = 0.001$, the correlation vector is defined as $z(5,\ 0.2)$ and the selection criteria is defined as $C = \{v_1, v_2 | v_1 * v_2 > 0 \& 0 \geq v_1 + v_2 \geq -1\}$. The upper bounds are set to only focus on small movements. We obtain Fig. 3.14 and We obtain Fig. 3.15.

Test 3: Sintel data-set; maximum displacement is $v_{max} = 38.6$. Initial values have a range of $v_0 = [0, -10]$, the magnitude constraint is $k = 0.001$, the correlation vector is defined as $z(5,\ 0.2)$ and the selection criteria is defined as $C = \{v_1, v_2 | v_1 * v_2 > 0 \& 0 \geq v_1 + v_2 \geq -10\}$. The upper bounds are set to only focus on small movements. We obtain Fig. 3.16 and Fig. 3.17.

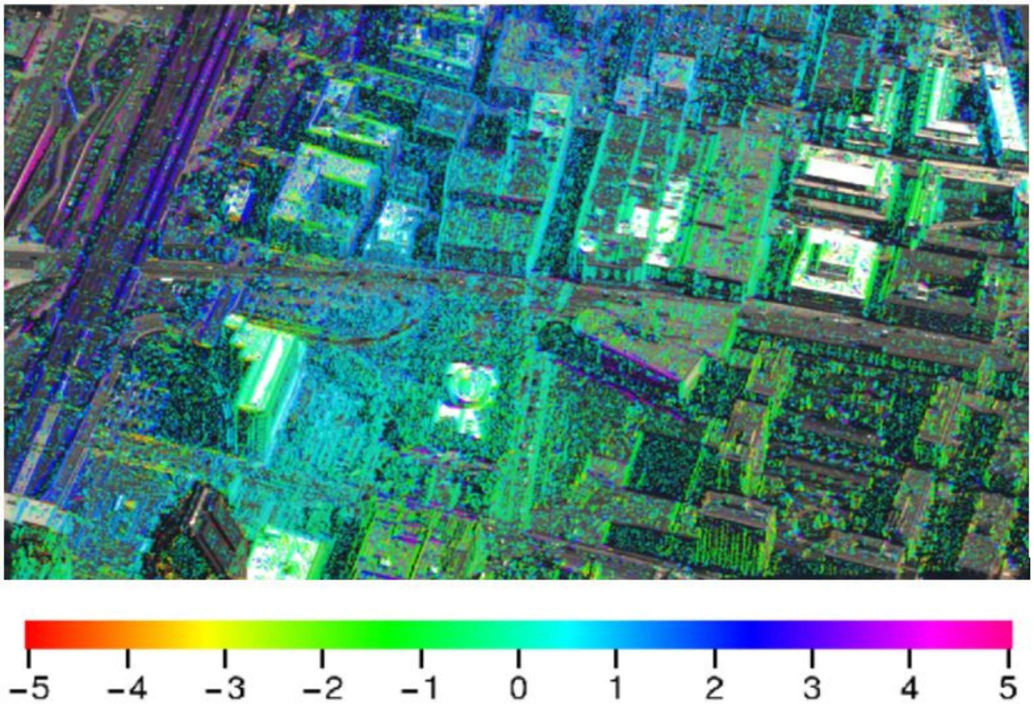


Figure 3.10: Full cyclopean solution for aerial imagery after rectification

Test 4: Sintel data-set $I_1(x)$ is artificially displaced $I_1(x - d * 0.1)$ where d is the ground truth; maximum displacement is $v_{max} = 3.86$. Initial values have a range of $v_0 = [0, -1]$, the magnitude constraint is $k = 0.0001$, the correlation vector is defined as $z(5,\ 0.2)$ and the selection criteria is defined as $C = \{v_1, v_2 | v_1 * v_2 > 0,\ 0 \geq v_1 + v_2 \geq -1\}$. The upper bounds are set to only focus on small movements. The data used in this test presents simulated noise such as: atmospheric effects, motion blur. focus blur, etc. We obtain Fig.3.18 and Fig. 3.19.

Test 5: Sintel data-set $I_2(x)$ is artificially displaced $I_2(x + d * 0.9)$ where d is the ground truth; maximum displacement is $v_{max} = 3.86$. Initial values have a range of $v_0 = [0, -1]$, the magnitude constraint is $k = 0.0001$, the correlation vector is defined as $z(5,\ 0.2)$ and the selection criteria is defined as $C = \{v_1, v_2 | v_1 * v_2 > 0 \& 0 \geq v_1 + v_2 \geq -1\}$. The upper bounds are set to only focus on small movements. The data used in this test

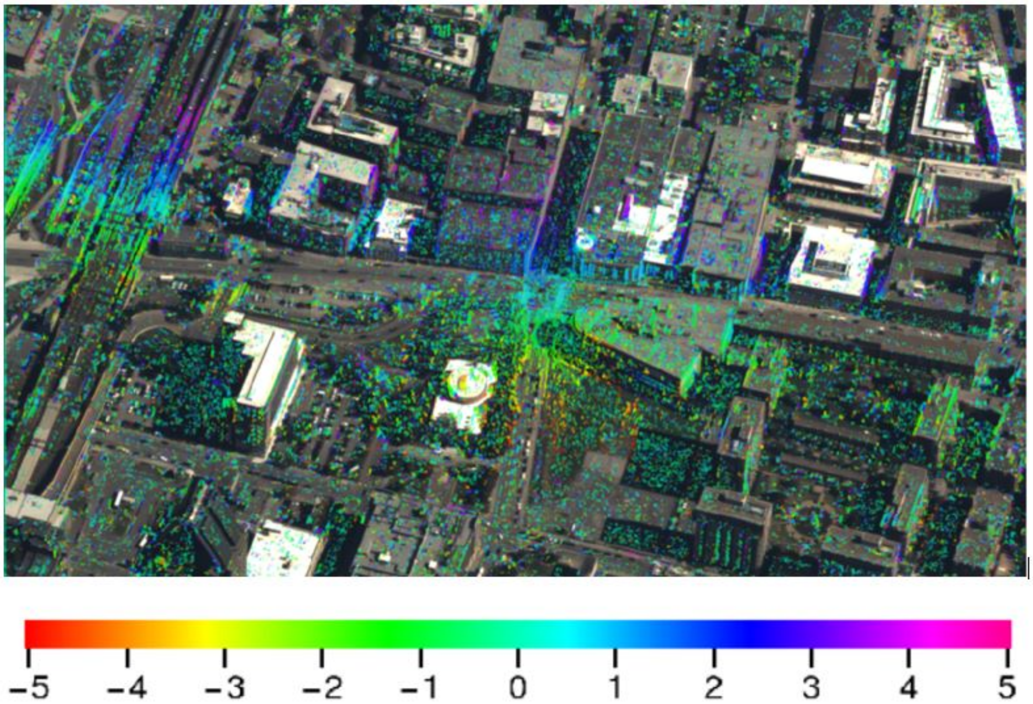


Figure 3.11: Full cyclopean solution for aerial imagery without rectification

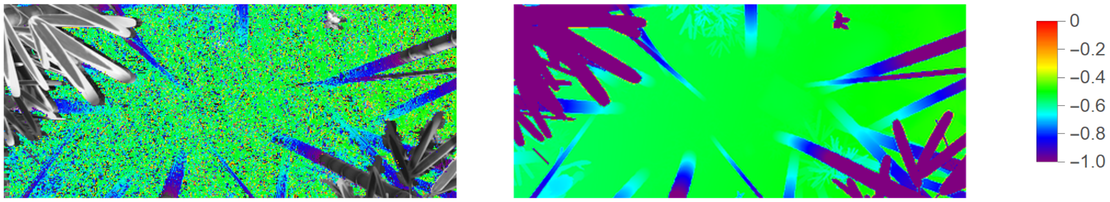


Figure 3.12: Test 1. Left: cyclopean optical flow, Right: ground truth.

presents simulated noise such as: atmospheric effects, motion blur. focus blur, etc. We obtain Fig.3.20 and Fig. 3.21.

Test 6: Sintel data-set; maximum displacement is $v_{max} = 38.6$. Initial values have a range of $v_0 = [0, -10]$, the magnitude constraint is $k = 0.0001$, the correlation vector

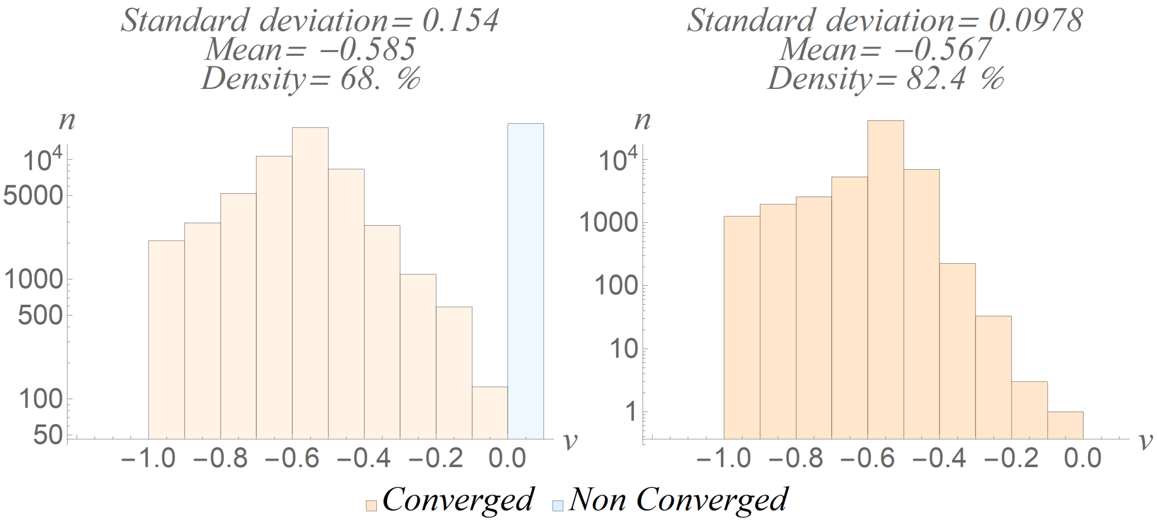


Figure 3.13: Test 1. Left: cyclopean optical flow histogram, Right: ground truth histogram.

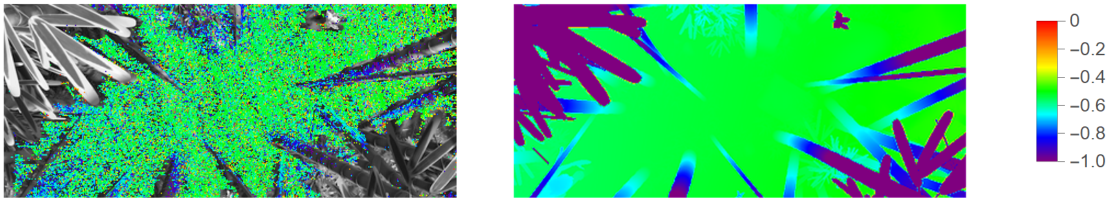


Figure 3.14: Test 2. Left: cyclopean optical flow, Right: ground truth.

is defined as $z(5,\ 0.2)$ and the selection criteria is defined as $C = \{v_1, v_2 | v_1 * v_2 > 0 \& 0 \geq v_1 + v_2 \geq -10\}$. The upper bounds are set to only focus on small movements. The data used in this test presents simulated noise such as: atmospheric effects, motion blur. focus blur, etc. We obtain Fig. 3.22 and Fig. 3.23.

It's possible to see from Table 3.1 that Test 1 haves the best RMSE score, this is because the optical flow is solved with *brightness consistency* when solving for $I_1(x)$ and

Table 3.1: Different methods solving optical flow.

Method	RMSE	Density
Test 1		
Cy	0.14	68%
$H\&S$	1.13	100%
LK	1.12	100%
Test 2		
Cy	0.17	52.1%
$H\&S$	1.15	100%
LK	1.344	100%
Test 3		
Cy	2.9	30.1%
$H\&S$	11.64	100%
LK	11.45	100%
Test 4		
Cy	0.13	72%
$H\&S$	1.13	100%
LK	1.12	100%
Test 5		
Cy	0.181	54.3%
$H\&S$	1.14	100%
LK	1.12	100%
Test 6		
Cy	2.8	31.4%
$H\&S$	11.61	100%
LK	11.45	100%

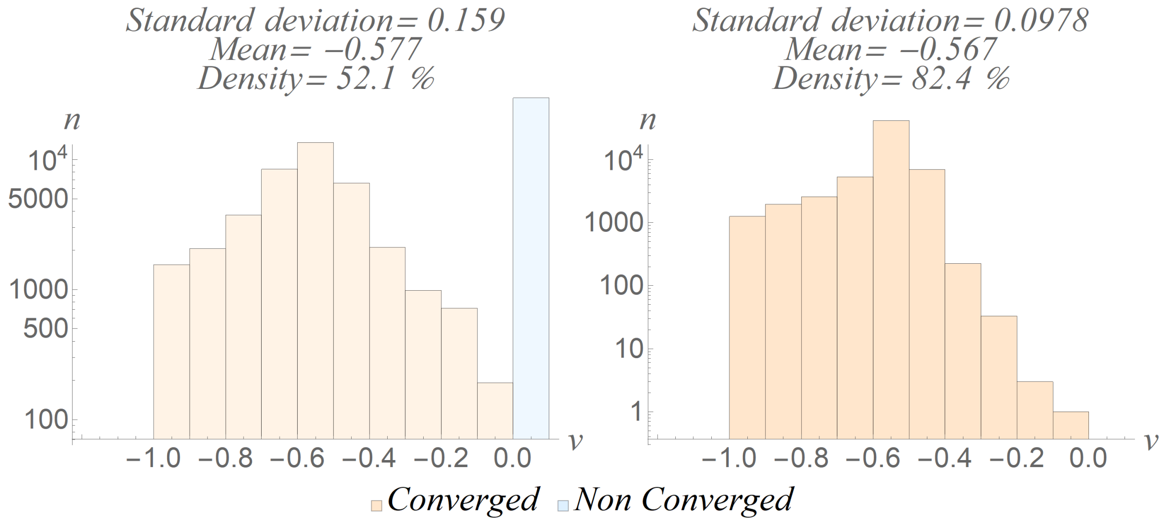


Figure 3.15: Test 2. Left: cyclopean optical flow histogram, Right: ground truth histogram.

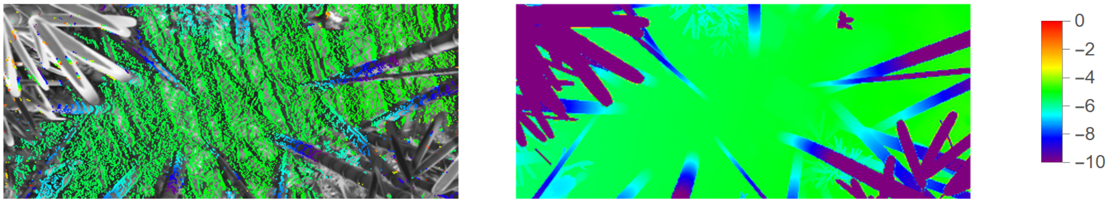


Figure 3.16: Test 3. Left: cyclopean optical flow, Right: ground truth.

$I_1(x - d * 0.1)$. Test 2, although artificial, is more reliable than Test 1; here we don't enforce *brightness consistency* because we solve for different intensity functions $I_1(x)$ and $I_2(x + d * 0.9)$. Test 3 does not have modified intensity functions, so it means that we are solving for big movements; as mentioned previously, big movements will lead to a deficient translation of C, which explains why Test 3 presents the lowest convergence density.

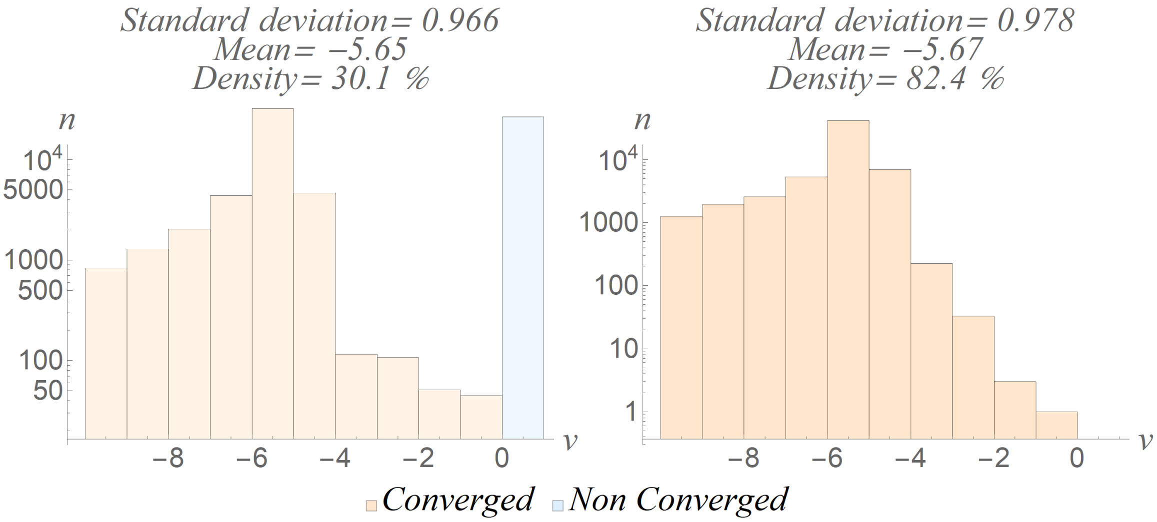


Figure 3.17: Test 3. Left: cyclopean optical flow histogram, Right: ground truth histogram.

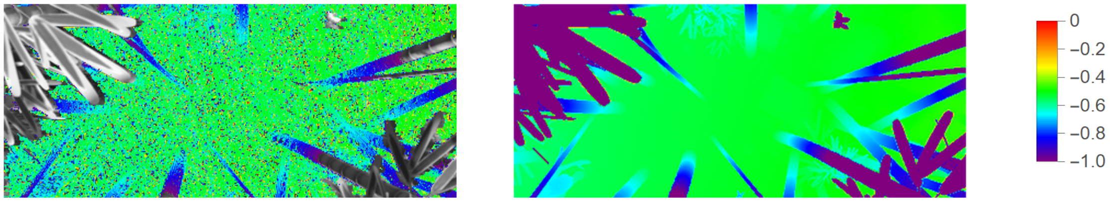


Figure 3.18: Test 4. Left: cyclopean optical flow, Right: ground truth.

Noise was also introduced for Test 1, 2 and 3 with the implementations of Test 4, 5 and 6 respectively. The obtained results show certain robustness in the method having variations smaller than ±0.1 for RMSE and ±0.4 for density.

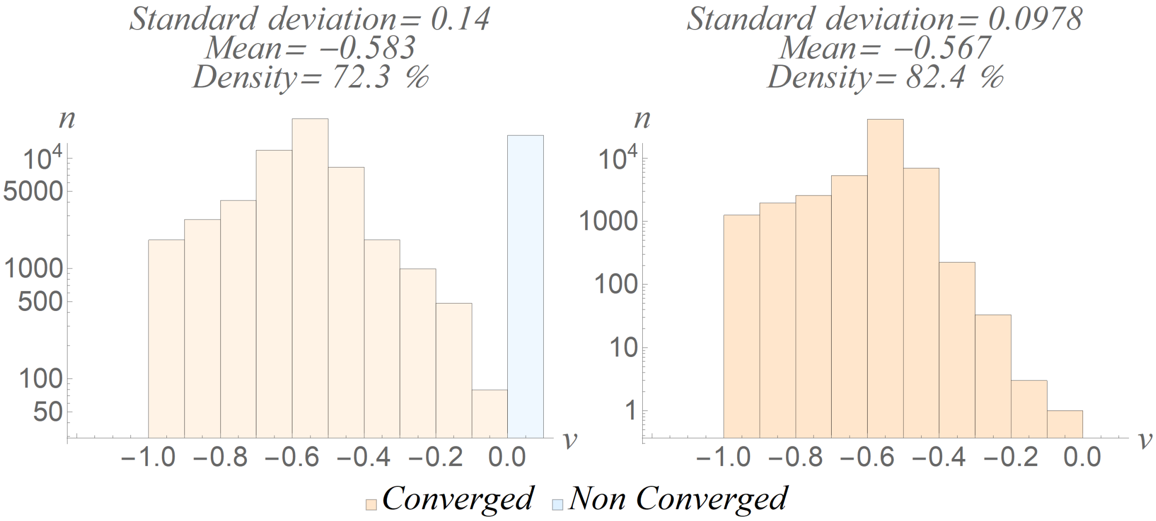


Figure 3.19: Test 4. Left: cyclopean optical flow histogram, Right: ground truth histogram.

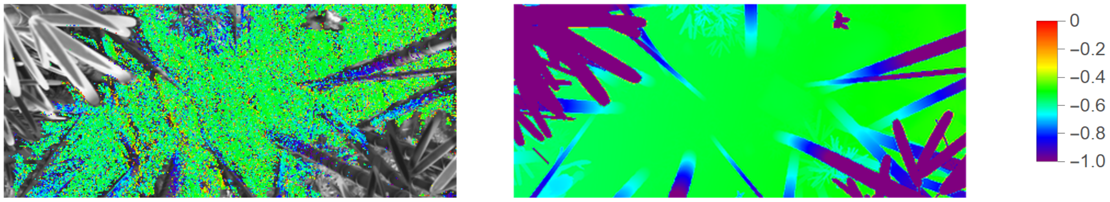


Figure 3.20: Test 5. Left: cyclopean optical flow, Right: ground truth.

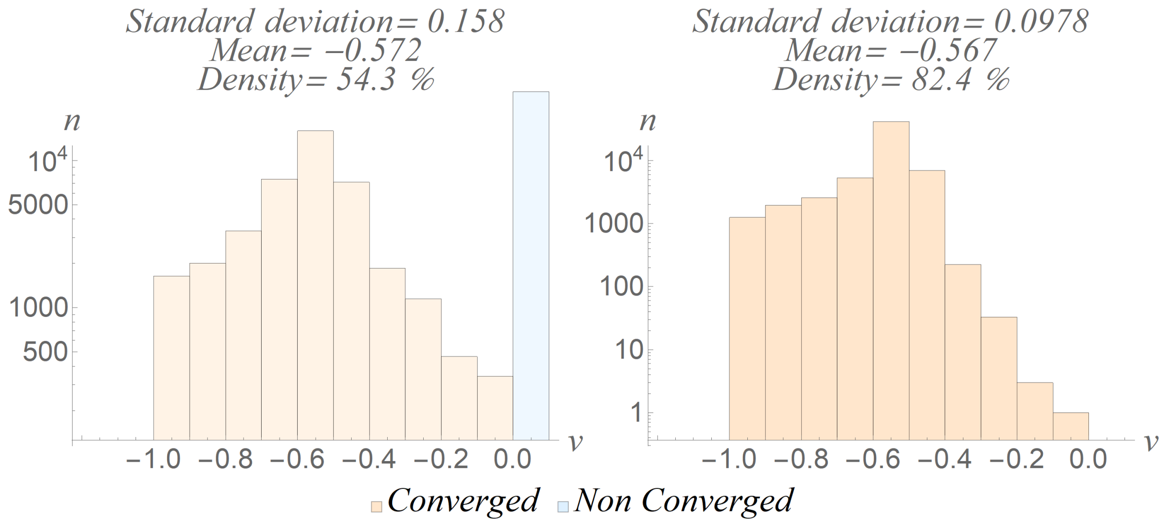


Figure 3.21: Test 5. Left: cyclopean optical flow histogram, Right: ground truth histogram.

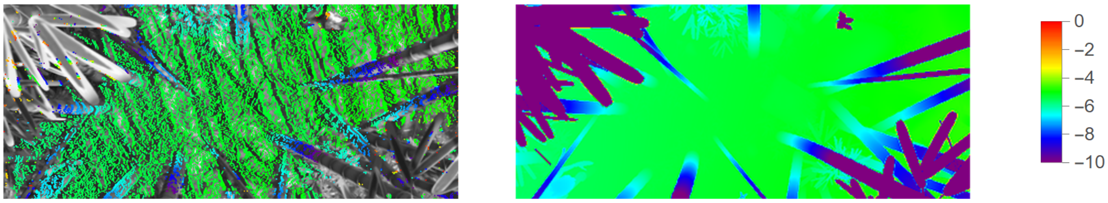


Figure 3.22: Test 6. Left: cyclopean optical flow, Right: ground truth.

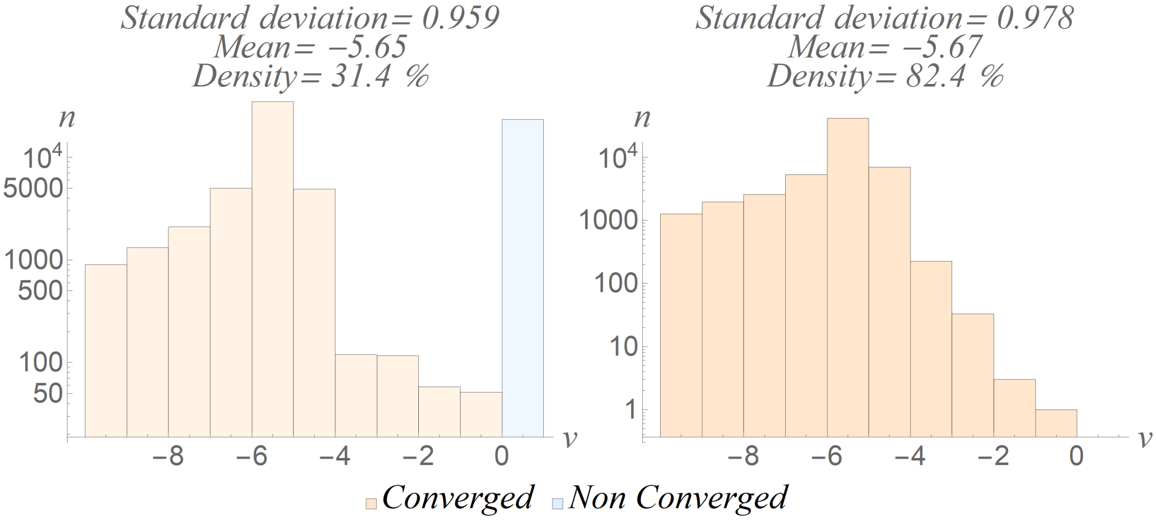


Figure 3.23: Test 6. Left: cyclopean optical flow histogram, Right: ground truth histogram.

Chapter 4

PYRAMIDS IN CYCLOPEAN OPTICAL FLOW

In practice, scenarios in images are far from ideal and introduce physical information this method won't consider; accordingly, due to the lack of capacity, sometimes it's not possible to interpret a solution. As a first assumption, a good solution won't present any of the behaviors discussed in section 2.2.3 because it's presumed to be an incorrect estimate.

If any gradient breaks the ideal behavior, the optical flow will be considered undefined.

Analysis of failure helps understand the limitations from a more affordable perspective. If an "undefined" state appears, we can find other means to estimate a solution.

The aperture problem is not only affecting 2D optical flow. In 1D situations, a lack of image texture will induce the same effect, since locally it will be impossible to estimate motion. It's important to mention, that 1D optical flow is directly solved when the gradient constraints are met. In 2D, the problem is under-constrained, so we must rely on extra information.

The scale of I will be modified in an attempt to find good results. The motivation behind this decision is that a simpler form of I will be more stable due to its smoother shape. Also, the construction of multiple images at various scale levels combines the information of a larger spatial neighborhood. We proposed to use a pyramidal representation I^n where n is the present level of resolution, more discussion about the subject in section 4.1.

Consider an example of a step edge between two uniform intensity regions, moving at a constant speed. As seen in fig 4.1, pixels $x = 6$ and $x = 15$ satisfies the gradient constraints and it's possible to find an estimate of motion. The remaining pixels' motion don't satisfy the gradient constraints therefore their motion is not defined. For the undefined pixels, the pyramid approach will alter the information of $I(x)$ and $I'(x)$ with the goal of supporting spatially zones that show an absence of information and achieving the gradient constraint.

The scale will provide a new intensity function that satisfies the gradient constraints, so the flow can be estimated at each pixel from x=1 to x=5.

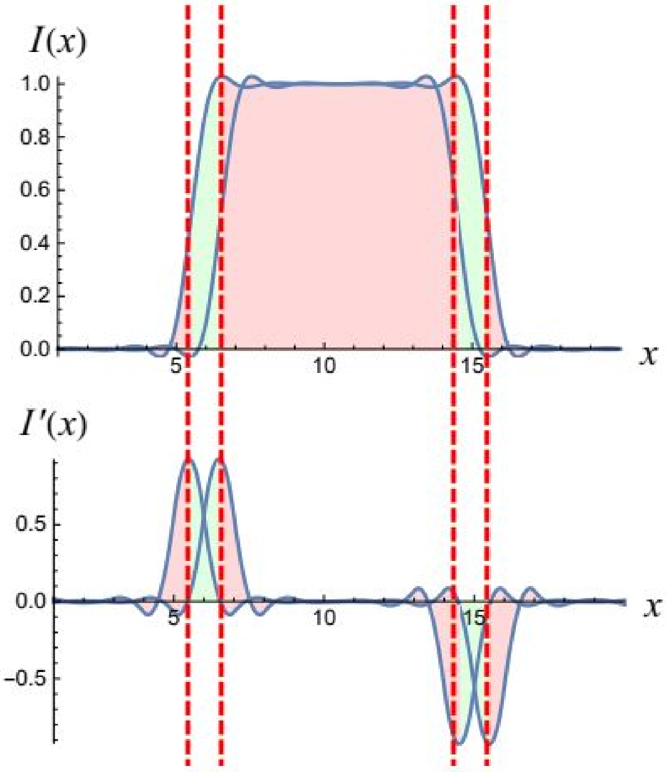


Figure 4.1

4.1 Scale

Reducing the scale of an image has been used often in computer vision to alter information with the goal of obtaining a preferable version of the image's signal. A pyramid is a multilevel image representation that computes n number of sub-samplings of the reference.

It's fairly common to divide size by a factor of 2 each time reduction is applied, therefore, in the algorithm proposed a down-sampling value of 2 is fixed.

To avoid loss of information, the blur effect is applied to load individual pixels of adjacent information. The usual method for blurring computes the equivalent of a pixel with a weighted window. The end result of moving the window over every pixel is a smooth version of the input signal.

With fitting values of weights, the process of smoothing will filter high-frequency signals and sudden changes in intensity. Thanks to this low pass filter behavior, the correlation between pixels is highlighted and noise discontinuities that might deform the scene are restrained.

Gaussian blur uses a Gaussian function to define the weight values to smoothen a signal, more details about it in section 4.1.1. The Gaussian behavior will manage importance in a radial form, being the center the most significant, which means the result will pay a lot of respect to the pixel's original form. This trait is important to preserve the essence of the signal's shape after each reduction, that's why it's proposed to use a Gaussian Pyramid for multilevel representations.

4.1.1 Gaussian pyramid

For a one dimensional blur, each pixel is convolved with a 5-pixel kernel ("5 tap" filter), which is defined as:

$$\hat{w} = \begin{pmatrix} c \\ b \\ a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \frac{1}{4} - \frac{a}{2} \\ \frac{1}{4} \\ a \\ \frac{1}{4} \\ \frac{1}{4} - \frac{a}{2} \end{pmatrix}$$

The value of a is a damping term linked to the correlation between pixels after blurring. Where $a = 0.5$ will minimize damping and the influence from the window's borders to zero.

Blurring an image $\ddot{I}$ takes the following computation:

$$\hat{\ddot{I}}(i) = \sum_{j=-2}^{2} \hat{w}(j) \cdot \ddot{I}(i-j) \tag{4.1}$$

It's important to mention that padding of size 2 is added to fill missing information required by the window when convolving in the borders of a signal.

After convolution, the result has a different distribution where individual pixels hold more spatial information but their capacity for describing details is lost. In accordance to downsampling, $\hat{\ddot{I}}$ will be subjected to $\ddot{J}(i) = \hat{\ddot{I}}(2i)$ to smoothing the signal even further and obtaining a sparse result ($\ddot{J}$).

Figure 4.2 illustrates the pyramid representation of $I(x)$.

Coming back to the example of fig 4.1, only pixels $x = 6$ and $x = 15$ could satisfy the gradient constraints, leaving the remaining pixels unsolved. But as seen in figure 4.3, altered forms of I meet the gradient constraints in a broader manner. For example, I^2 will solve optical flow for every pixel except $x = 10$ and $x = 11$. It's important to emphasize that is possible to overscale a signal, in this example, I^3 is meeting a downgrade on the convergence domain where $x = 9$ to $x = 12$ does not satisfy the gradient constraints.

4.1.2 Solving optical flow over multiple scales

The iterative solving equations (2.4) and (2.5) will be referred to as the function (4.2) where v_0 is the initial value of v.

$$v = upgrade(v_0) \tag{4.2}$$

This definition only considers the base level of an image's pyramid. To solve optical

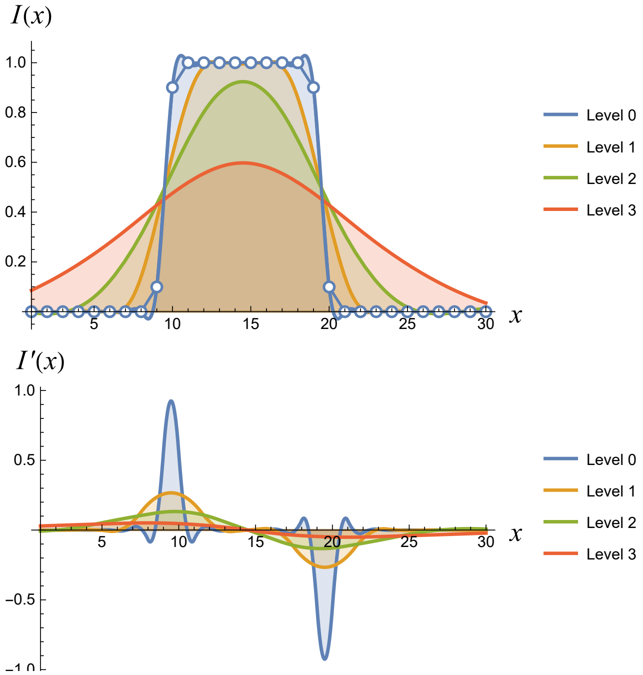


Figure 4.2: Original signal $I(x)$ and it' 3 subsequent levels of a pyramidal representation.

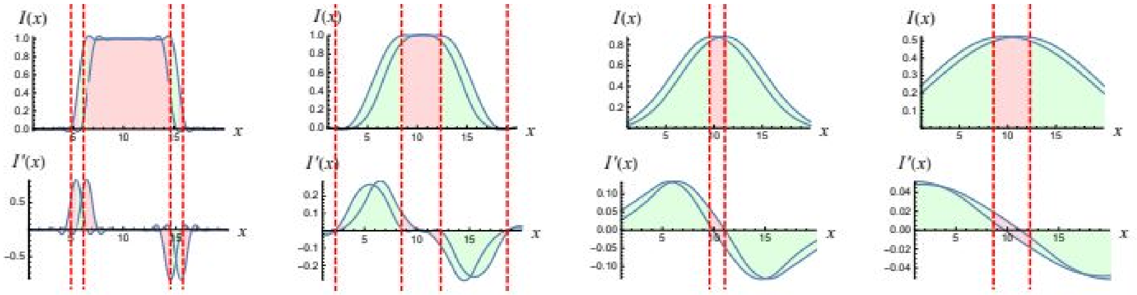


Figure 4.3: Most left: convergence of base level of I's pyramid. Most right: convergence of highest level of I's pyramid.

flow over multiple scales, the same upgrade function is applied updating I^{n-1} after solving v^n for I^n.

The iteration starts at the smallest scale (top of the pyramid) and is upgraded as follows:

$$
\begin{aligned}
v^1 &= upgrade(v_0) \\
v^2 &= upgrade(v^1) \\
&\dots \\
v^{n+1} &= upgrade(v^n)
\end{aligned}
\tag{4.3}
$$

At each level, if the solution does not satisfy the constraints (sections 2.2.3 and 2.2.4), we stop.

As seen in fig. 4.3, the base level of an interpolated step edge generally won't meet the gradient constraints, but it'll have a wider reach when explored in different resolutions.

The base level of an image's pyramid usually contains more details and noise compared to its higher levels, these characteristics are not ideal for optical flow estimation due to the discontinuities it produces in the spatial derivatives; however, during the process of sub-sampling, signals will smoothly be altered until no longer precise. These limitations motivate the algorithm to iterate over every pyramid level from the lowest resolution I^n to I^0, with the objective of estimating v^0 eluding the constraints by referencing previous solutions of v.

In fig 4.4, the only analyzed signal is I^0, where the only possible solution is $v = undefined$ because the gradient doesn't meet the constraints. A non-pyramidal representation would then not have enough capacity to solve this example.

4.2 Initial Values

During the process of using the update function in 2.1, a convergence map can be visualized if the input values of v are used as a two component argument. In 2.2.2 is identifiable that convergence also depends on the possible combinations of v_1 and v_2 to form v for a given p. This characteristic is then used to increase the robustness of the

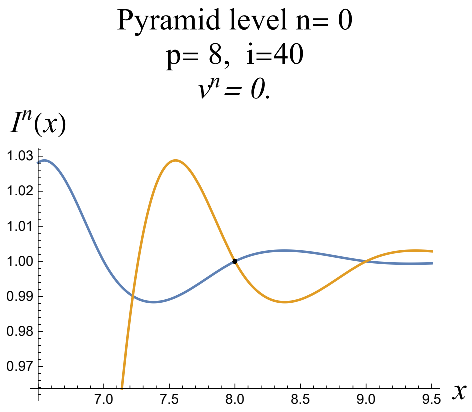


Figure 4.4

method, by relying on the capacity of v to be arbitrary. Any pyramidal upgrade 4.4 at level $n+1$ will be initialized with a list $V^n := \{\{v_1, v_2\} | v_1 + v_2 = v^n, v_1 \times v_2 \geq 0\}$ where input is the coerced solution from the previous level n,

$$\begin{aligned} v^1 &= upgrade(V_0) \\ v^2 &= upgrade(V^1) \\ &\ldots \\ v^{n+1} &= upgrade(V^n) \end{aligned} \tag{4.4}$$

The amount of levels used to solve a given stereo problem depends on the texture, image dimensions and maximum displacement estimated. If we consider our images highly detailed and with big displacements, it should be represented with several pyramidal levels.

4.3 Pyramidal Constraints

During the calculations of optical flow, we might encounter constraints that lead to undefined values (subsections 2.2.3 and 2.2.4). In pyramidal optical flow, we transmit the undefined values of v to lower levels of the pyramid, rejecting any future update of

v.

In this section we propose to handle the pyramid's constraint propagation in a different manner: During the pyramidal update, we will allow a e amount of errors, where $0 \geq e \geq n$, being n the pyramid's highest level; $e = n$ will broadly apply the first encountered constraint, $e = 0$ will propagate none.

Intermediate values of e considers as a solution any level's converged value unless a flow converges in lower levels of the pyramid, where data is more faithful to the original signal.

This process is different than other methods because we posses more information due to the cyclopean nature of the vectors. They will provide capacity to describe a solution with the addition of constraining it as well.

4.4 Experiments

4.4.1 Synthetic Validation of pyramids

We will take as an example the pyramidal representation in figure 4.2 and solve for $p = 13$ to analyze the convergence zone behavior when transferring optical flow along levels.

For the experiments we consider:

$$\begin{aligned}
&C = \{v_1, v_2 |\ v_1 * v_2 > 0\ \&\ 0 \geq v_1 + v_2 \geq 2\} \\
&k = 0.001 \\
&v^0 \in \mathbb{R} : -3 \leq v^0 \leq 3 \\
&z(n, \Delta n) = z(2, 0.2) \\
&e = 0
\end{aligned} \tag{4.5}$$

When we refer to the results on the Fig. 2.11, it's noticeable that the convergence zone becomes larger through the addition of pyramidal levels, as seen in fig. 4.5. With this behavior, solutions are accessed more robustly, therefore the list of initial values could become smaller having a faster method with the coarse-to-fine approach.

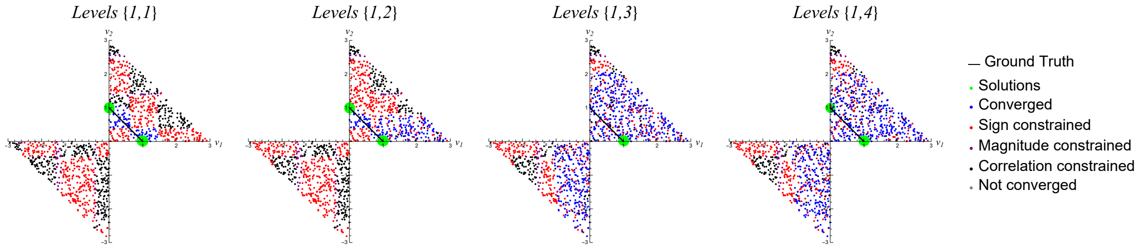


Figure 4.5: Convergence map of signal for different pyramidal descents when $p = 13$.

The pyramidal cyclopean method propagates optical flow through a pyramidal representation. In the level 3, we manage to find the solution $v_1, v_2 = 0.38, 0.47$ With an initial value of $v_0 = [0, 0]$; The solution will be transferred to the next level as initial value. This process continues until we arrive to the base of the pyramid with the solution $\{v_1, v_2\} = \{1.0034, 0.025\}$, as seen in figure 4.6.

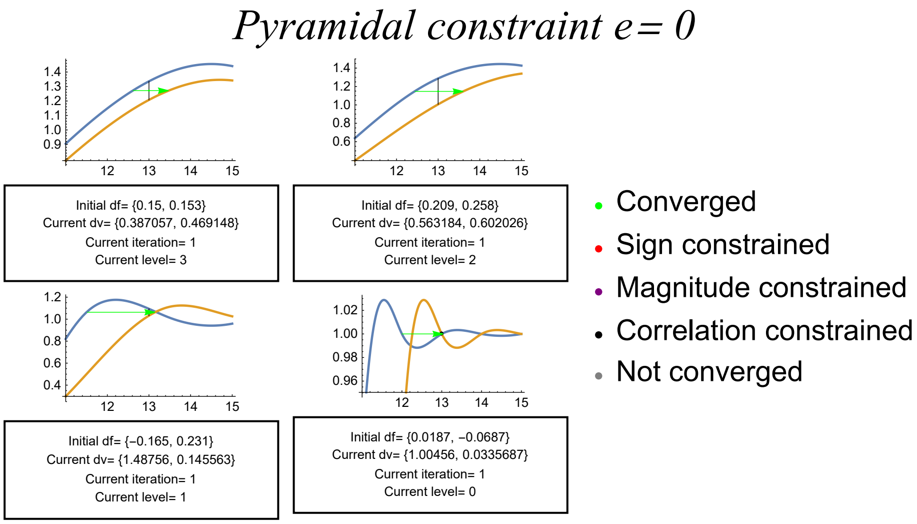


Figure 4.6: Convergence for pixel $p = 13$ in different levels.

More solutions are presented in Fig. 4.7. It's important to mention that less points will converge in comparison of the original cyclopean application (as seen in Fig. 2.15) due to the pyramidal over-constraint $e = 0$, where constraints encountered along the levels will be propagated up to the base of the pyramid as seen in section 4.3. For example, the method fails in $p = 8$, along the pyramids; As seen in Fig. 4.8 $p = 8$ is updated according to the pyramid's levels. The solution from level 3 is updated every time a better solution is found, this process happens until level 0. At level 0, the shape of the curve suffers from noise and interpolation inaccuracy, making it not reach the gradient standards for analysis. Therefore the solution $v^0 = undefined$ and is rejected to become $v^0 = v^1$.

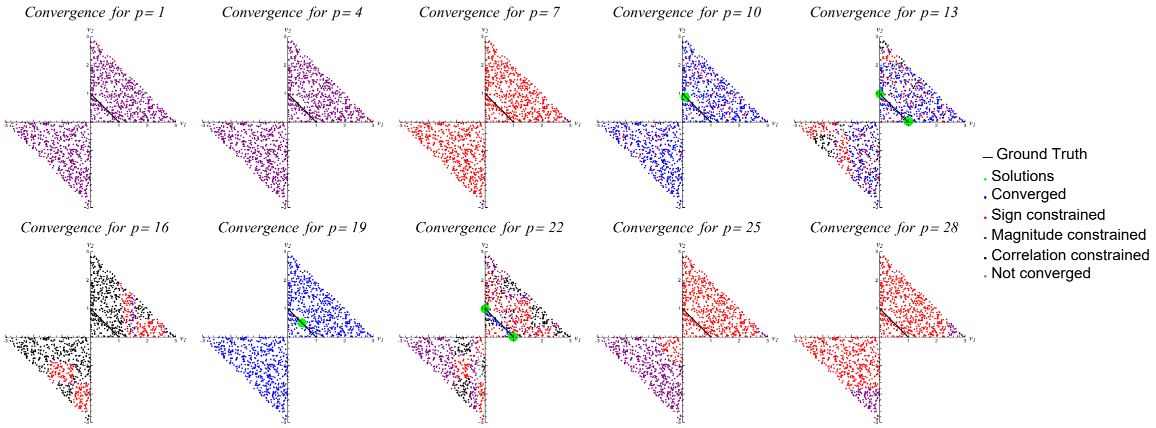


Figure 4.7: Convergence for different points in signal.

We add ±0.05 noise to our signals solving for different points as seen in fig. 4.11. It's possible to notice the increase in uniformity in points like $p = 19$ and 10 in comparison to the performance of the cyclopean method over noise in fig. 2.17.

Synthetic validation of pyramidal constraint

No constraints ($e = n$) means every point haves a solution, being the last accepted value always propagated through the pyramidal levels as seen in Fig.4.11. The value of e can be modified to increase density but loosing precision.

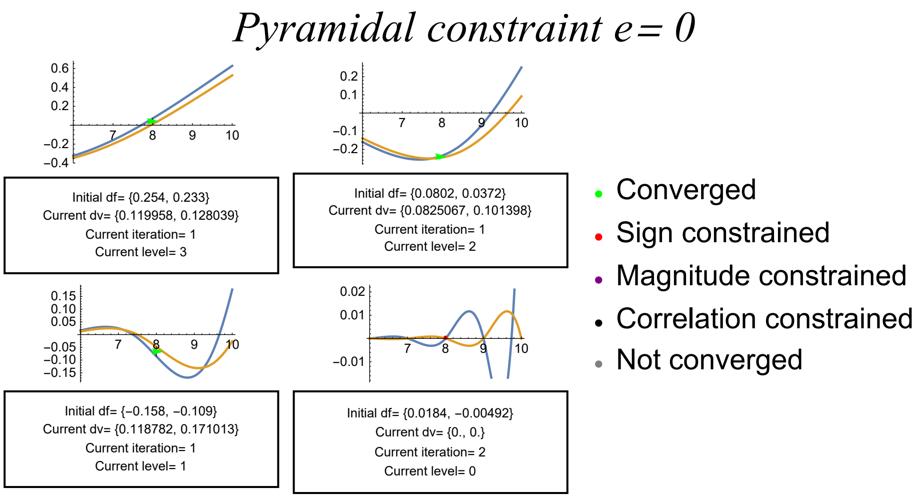


Figure 4.8: Propagation of constraint in $p = 8$.

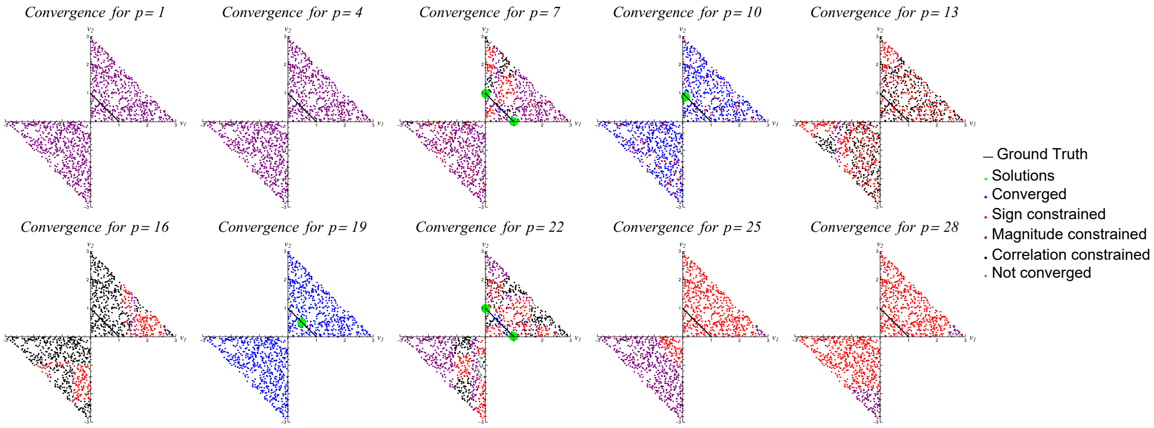


Figure 4.9: Convergence in different levels when the signals present noise.

4.4.2 Stereo results

Test 0: Aerial imagery

We replicate the experiment in subsection 3.3.1 with the addition of pyramidal propagation through the levels $\{1, 4\}$.

When we solve pyramidal optical flow for Fig. 1.2 we obtain Fig. 4.12. The ground truth is unknown, but it's possible to use the context in the image to conclude the

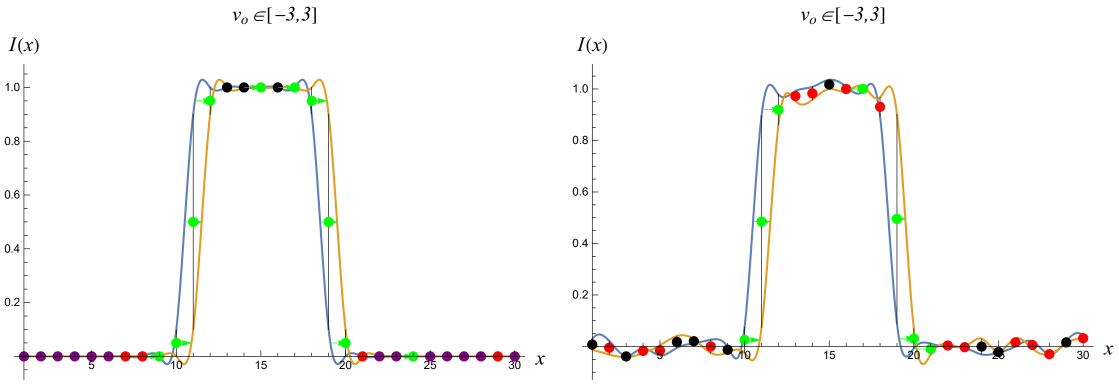


Figure 4.10: Convergence of signal, Left: Pure signal, Right: Signal with noise.

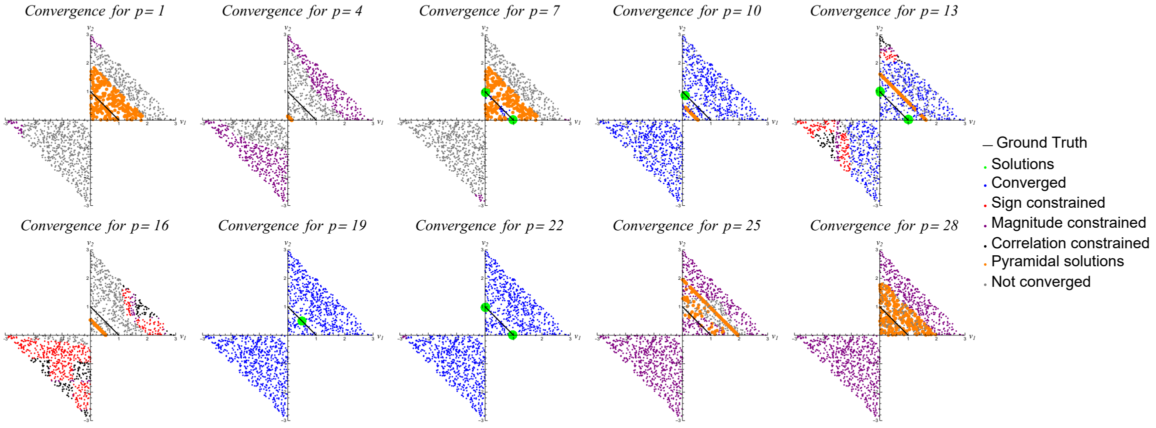


Figure 4.11: Convergence in different levels when the pyramidal constraint $e = n$.

performance in this circumstance: It's possible to see the top of the buildings will tend to have a disparity higher than 0.6, on the other hand, anything surpassing the ceiling section looses magnitude. This respects the physical property of perspective, where far away objects move slower than closer ones, for example, the ceilings in the image. It's

Figure 4.12: Cyclopean optical flow solution of aerial imagery

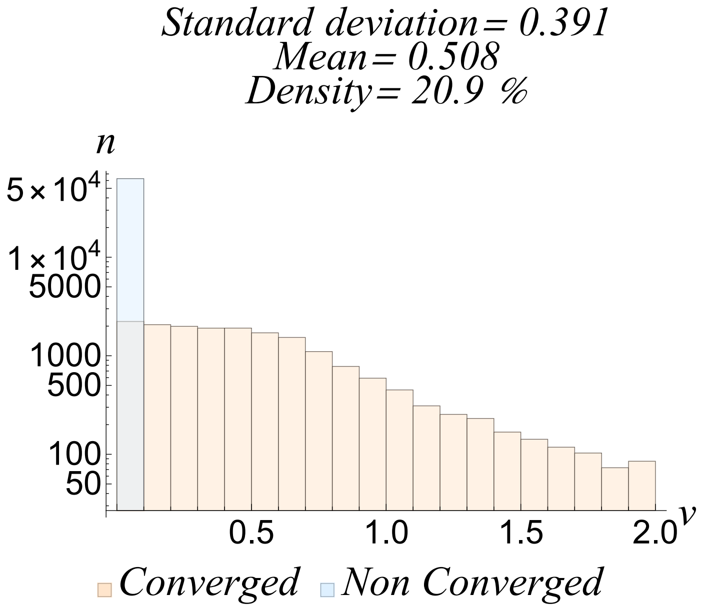


Figure 4.13: Cyclopean optical flow solution of aerial imagery

possible to see also that the lack of texture is affecting the density, with only 20.9% pixels of the image converged.

In this section we obtain more uniform results (lower standard deviation) but a decrease in density, in comparison to the previous Test 0 in subsection 3.3.1. In Fig.4.14 we implement $e = 2$ to allow certain degree of approximation. Density will increase to 25.3% and still present the same distribution as previous tests.

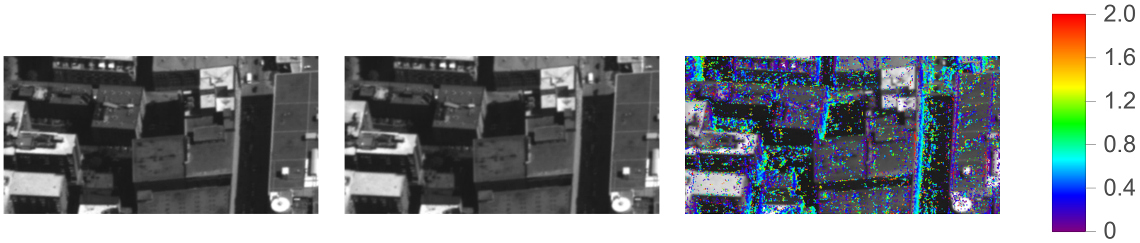


Figure 4.14: Cyclopean optical flow solution of aerial imagery

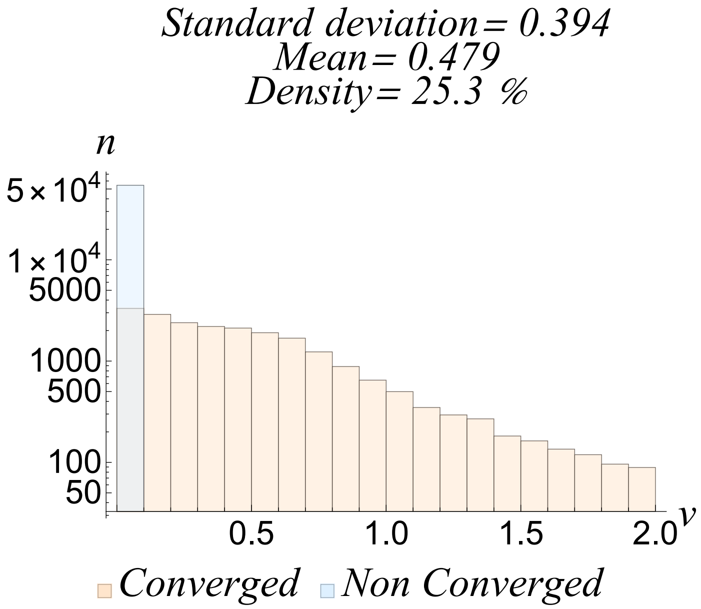


Figure 4.15: Cyclopean optical flow solution of aerial imagery

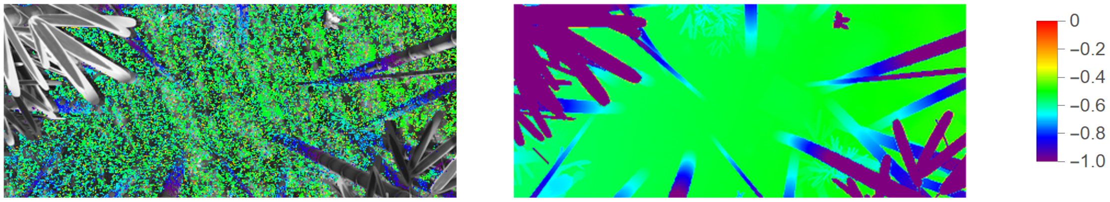


Figure 4.16: Test 1. Left: cyclopean optical flow, Right: ground truth.

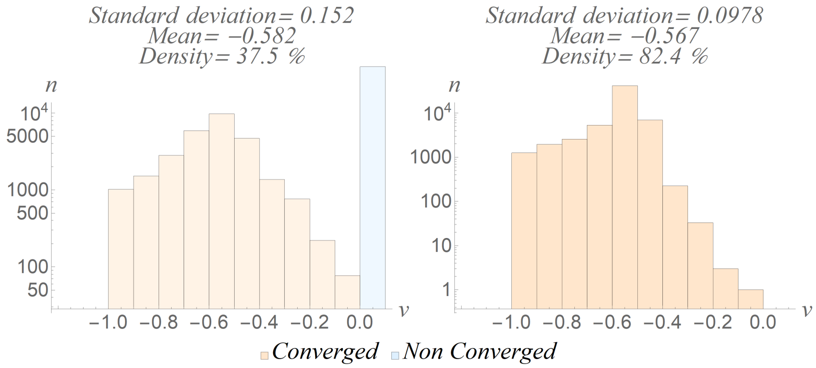


Figure 4.17: Test 1. Left: cyclopean optical flow histogram, Right: ground truth.histogram

Sintel data-set

We replicate the experiment in subsection 3.3.2 with the addition of pyramidal propagation through levels $\{1, 4\}$.

Test 1: Sintel data-set $I_1(x)$ is artificially displaced $I_1(x - d * 0.1)$ where d is the ground truth. We obtain Fig. 4.16 and Fig. 4.17.

Test 2: Sintel data-set $I_2(x)$ is artificially displaced $I_1(x + d * 0.9)$ where z is the ground truth. We obtain Fig.4.18 and Fig.4.19.

Test 3: Sintel data-set. We obtain Fig. 4.20 and Fig. 4.21.

Test 4: Sintel data-set $I_1(x)$ is artificially displaced $I_1(x - d * 0.1)$ where d is the

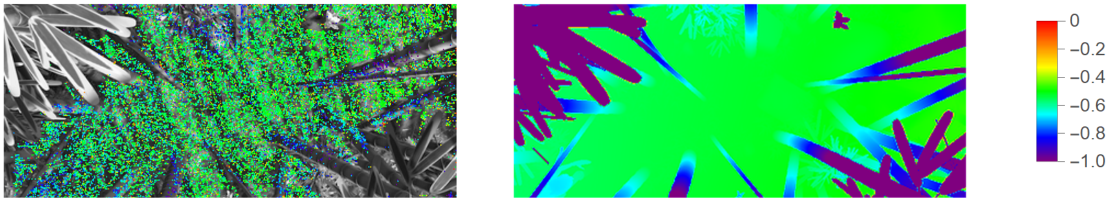


Figure 4.18: Test 2. Left: cyclopean optical flow, Right: ground truth.

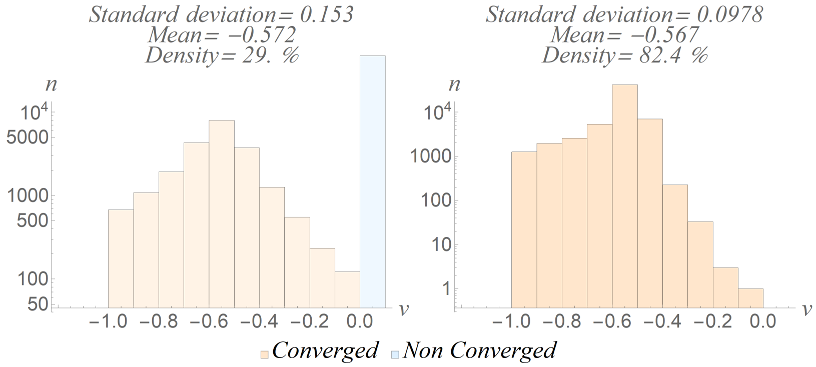


Figure 4.19: Test 2. Left: cyclopean optical flow histogram, Right: ground truth histogram

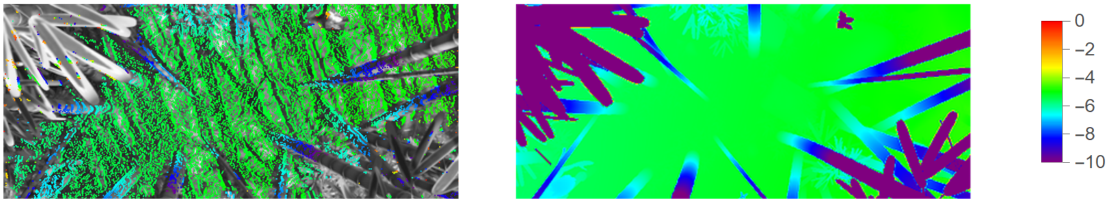


Figure 4.20: Test 3. Left: cyclopean optical flow, Right: ground truth.

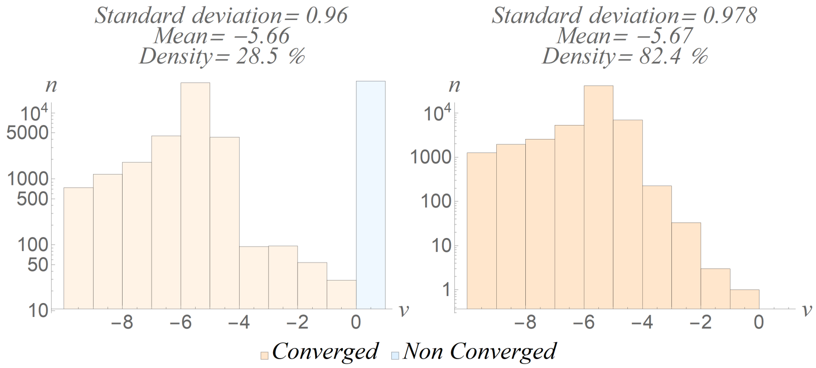


Figure 4.21: Test 3. Left: cyclopean optical flow histogram, Right: ground truth histogram

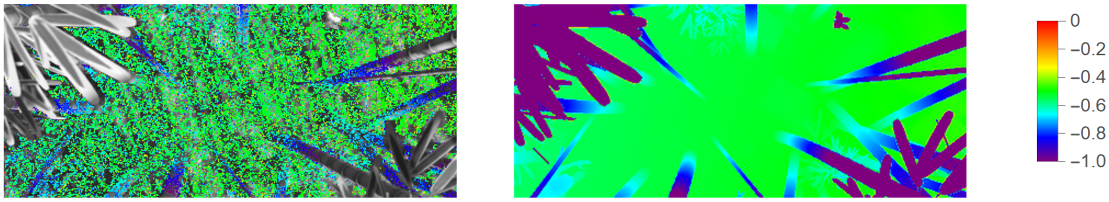


Figure 4.22: Test 4. Left: cyclopean optical flow, Right: ground truth.

ground truth. The data used in this test presents simulated noise such as: atmospheric effects, motion blur. We obtain Fig. 4.22 and Fig. 4.23.

Test 5: Sintel data-set $I_1(x)$ is artificially displaced $I_1(x + d * 0.9)$ where d is the ground truth. The data used in this test presents simulated noise such as: atmospheric effects, motion blur. We obtain Fig. 4.24 and Fig. 4.25.

Test 6: Sintel data-set. The data used in this test presents simulated noise such as: atmospheric effects, motion blur. We obtain Fig. 4.26 and Fig. 4.27.

It's possible to see from Table 4.1 that Test 1 has the best numbers, this is because the optical flow is solved with *brightness consistency* when solving for $I_1(x)$ and $I_1(x -$

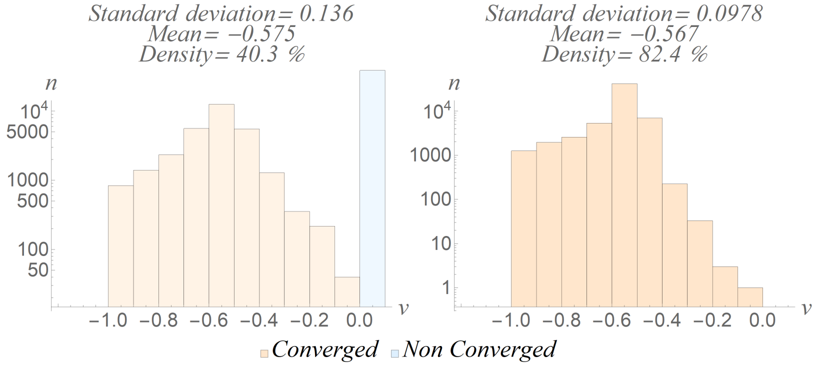


Figure 4.23: Test 4. Left: cyclopean optical flow histogram, Right: ground truth.histogram

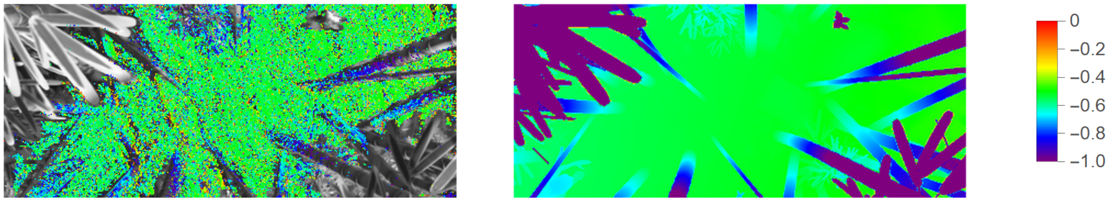


Figure 4.24: Test 5. Left: cyclopean optical flow, Right: ground truth.

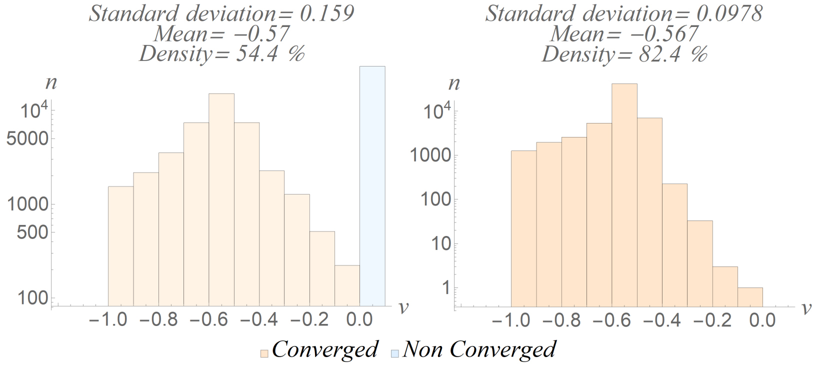


Figure 4.25: Test 5. Left: cyclopean optical flow histogram, Right: ground truth.histogram

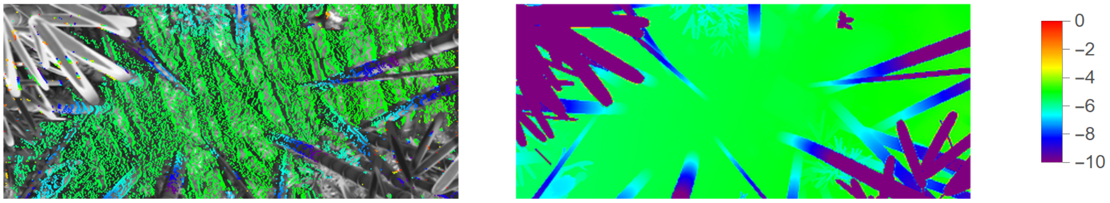


Figure 4.26: Test 6. Left: cyclopean optical flow, Right: ground truth.

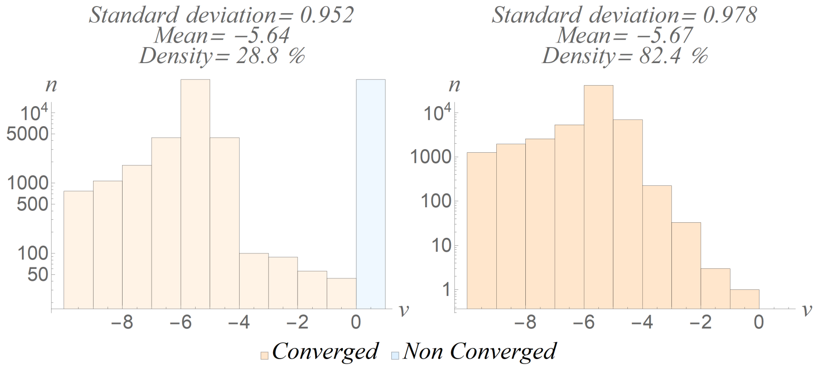


Figure 4.27: Test 6. Left: cyclopean optical flow histogram, Right: ground truth.histogram

Table 4.1: Different tests solving optical flow

Test	RMSE	Density
Test 1	0.1397	37.5%
Test 2	0.164	29.7%
Test 3	2.9	28.6%
Test 4	0.124	40.3%
Test 5	0.182	54.4%
Test 6	5.49621	28.915%

$z * 0.1$). Test 2, although artificial, is more reliable than Test 1; here we don't enforce *brightness consistency* because we solve for different intensity functions solving for $I_1(x)$ and $I_2(x + z * 0.9)$. Test 3 does not have modified intensity functions, so it means that we are solving for big movements; as mentioned previously, big movements will lead to a deficient translation of C, which explains why Test 3 presents the lowest convergence density. If we compare these results to Table 3.1, we can see the pyramidal application will greatly over-constrain density for small movements, but this doesn't transfer to big movements like Test 3 and 6.

CONCLUSION

Cyclopean optical flow achieved comprehensive results in a really clean and controlled matter. We can observe optical flow does solve stereo matching. Small motion and ideal brightness are beneficial for the method. An area of opportunity is to consider how to manage the lack of texture in aerial imagery and a future application of pyramidal representation. Stereo reconstruction and real-time depth estimation are also to be explored in future work.

Assuming the camera motion is small, we can try to estimate the camera geometry with 1D optical flow. This is not possible for the general camera motion cases, but it applies to the case of pure translational motion.

If parallel epipolar lines are assumed, we can find the single orientation of those lines to be able to rectify and then solve for stereo.

Between stereo pairs, if the dominant source of motion is from the parallax induced by depth variations, then the optical flow measured in the direction of the "true" epipolar lines should be larger than in any other direction.

Thus we propose to estimate 1D optical flow in many directions over the whole image and pick the one that results in the largest flow magnitude

Optical flow is a challenging field, and a constrained cyclopean approach is designed to take that into account. This allows the solution to only return confident results, being a future project to apply for this work on real-time depth estimation.

Some challenge of this work is that the pyramidal approach resulted inconclusive and therefore lack of capacity to describe large movements. It will be interesting to test the pyramidal under-constraint, to make a better prediction with the extracted estimations. More tests are necessary to explore further defects and advantages of the method.

BIBLIOGRAPHY

---

[1] Edward H Adelson. The extraction of spatio-temporal energy in human an machine vision. In *Proceedings from the workshop on motion representation and analysis, Charleston, SC, May 7-9ˆ< th>, 1986*, 1986.

[2] EH Adelson, EP Simoncelli, and William T Freeman. Pyramids and multiscale representations. *Representations and Vision, Gorea A.,(Ed.). Cambridge University Press, Cambridge*, pages 3–16, 1991.

[3] Min Bai, Wenjie Luo, Kaustav Kundu, and Raquel Urtasun. Deep semantic matching for optical flow. *CoRR*, abs/1604.01827, 2016. URL `http://arxiv.org/abs/1604.01827`.

[4] Simon Baker, Stefan Roth, Daniel Scharstein, Michael J. Black, J.P. Lewis, and Richard Szeliski. A database and evaluation methodology for optical flow. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, 2007. doi: 10.1109/ICCV.2007.4408903.

[5] John Barron, David Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12:43–77, 02 1994. doi: 10.1007/BF01420984.

[6] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11):1222–1239, 2001.

[7] Myron Z. Brown, Darius Burschka, and Gregory D. Hager. Advances in computational stereo. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 25(8):993–1008, 2003.

[8] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *European conference on computer vision*, pages 25–36. Springer, 2004.

[9] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision 61*, page 211–231, 2005.

[10] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley, and Michael J. Black. A naturalistic open source movie for optical flow evaluation. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision – ECCV 2012*, pages 611–625, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-33783-3.

[11] Jia-Ren Chang and Yong-Sheng Chen. Pyramid stereo matching network. *CoRR*, abs/1803.08669, 2018. URL `http://arxiv.org/abs/1803.08669`.

[12] Qifeng Chen and Vladlen Koltun. Full flow: Optical flow estimation by global optimization over regular grids. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4706–4714, 2016.

[13] Yang Chen, Martin Alain, and Aljosa Smolic. Fast and accurate optical flow based depth map estimation from light fields. *CoRR*, abs/2008.04673, 2020. URL `https://arxiv.org/abs/2008.04673`.

[14] Julie Delon and Bernard Rougé. Small Baseline Stereovision. *Journal of Mathematical Imaging and Vision*, 28(3):209–223, October 2007. doi: 10.1007/s10851-007-0001-1. URL `https://hal.archives-ouvertes.fr/hal-02269064`.

[15] Sotirios Ch. Diamantas, Anastasios Oikonomidis, and Richard M. Crowder. Depth computation using optical flow and least squares. In *2010 IEEE/SICE*

International Symposium on System Integration, pages 7–12, 2010. doi: 10.1109/SII.2010.5708293.

[16] Tiago J. Dias, Helder Araújo, and Pedro Miraldo. 3d reconstruction with low resolution, small baseline and high radial distortion stereo images. *CoRR*, abs/1709.06451, 2017. URL `http://arxiv.org/abs/1709.06451`.

[17] R. Feris, Ramesh Raskar, Longbin Chen, Kar-Han Tan, and Matthew Turk. Discontinuity preserving stereo with small baseline multi-flash illumination. volume 1, pages 412– 419 Vol. 1, 11 2005. ISBN 0-7695-2334-X. doi: 10.1109/ICCV.2005.76.

[18] Philipp Fischer, Alexey Dosovitskiy, Eddy Ilg, Philip Häusser, Caner Hazırbaş, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks, 2015.

[19] David J. Fleet and Yair Weiss. Optical flow estimation, 2005.

[20] D.J. Fleet and A.D. Jepson. Stability of phase information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(12):1253–1268, 1993. doi: 10.1109/34.250844.

[21] Denis Fortun, Patrick Bouthemy, and Charles Kervrann. Optical flow modeling and computation: A survey. *Computer Vision and Image Understanding*, 134:1–21, 2015. ISSN 1077-3142. doi: https://doi.org/10.1016/j.cviu.2015.02.008. URL `https://www.sciencedirect.com/science/article/pii/S107731421500`
Image Understanding for Real-world Distributed Video Networks.

[22] David Freedman, Robert Pisani, and Roger Purves. Statistics (international student edition). *Pisani, R. Purves, 4th edn. WW Norton & Company, New York*, 2007.

[23] David Gadot and Lior Wolf. Patchbatch: a batch augmented loss for optical flow, 2016.

[24] Fatma Güney and Andreas Geiger. Deep discrete flow. In *ACCV*, 2016.

[25] Xiaoyang Guo, Kai Yang, Wukui Yang, Xiaogang Wang, and Hongsheng Li. Group-wise correlation stereo network. *CoRR*, abs/1903.04025, 2019. URL `http://arxiv.org/abs/1903.04025`.

[26] Damith C. Herath, K. R. S. Kodagoda, and Gamini Dissanayake. Modeling errors in small baseline stereo for slam. In *2006 9th International Conference on Control, Automation, Robotics and Vision*, pages 1–6, 2006. doi: 10.1109/ICARCV.2006.345368.

[27] HD Herath, S Kodagoda, and G Dissanayake. *Stereo vision based SLAM: Issues and solutions*. ITECH, 2007.

[28] Heiko Hirschmüller and Daniel Scharstein. Evaluation of cost functions for stereo matching. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[29] Berthold K.P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1):185–203, 1981. ISSN 0004-3702. doi: https://doi.org/10.1016/0004-3702(81)90024-2. URL `https://www.sciencedirect.com/science/article/pii/00043702819002`

[30] Zhicong Huang, Xuemei Hu, Zhou Xue, Weizhu Xu, and Tao Yue. Fast light-field disparity estimation with multi-disparity-scale cost aggregation. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6300–6309, 2021. doi: 10.1109/ICCV48922.2021.00626.

[31] Junhwa Hur and Stefan Roth. Optical flow estimation in the deep learning age, 2020.

[32] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks, 2016.

[33] Saad Imran, Muhammad Umar Karim Khan, Sikander Mukaram, and Chong-Min Kyung. Unsupervised monocular depth estimation with multi-baseline stereo. In *BMVC*, 2020.

[34] Saad Imran, Sikander Bin Mukarram, Muhammad Umar Karim Khan, and Chong-Min Kyung. Unsupervised deep learning for depth estimation with offset pixels. *Opt. Express*, 28(6):8619–8639, Mar 2020. doi: 10.1364/OE.385328. URL `http://opg.optica.org/oe/abstract.cfm?URI=oe-28-6-8619`.

[35] Neel Joshi and Larry Zitnick. Micro-baseline stereo. Technical Report MSR-TR-2014-73, Microsoft, May 2014. URL `https://www.microsoft.com/en-us/research/publication/micro-basel`

[36] Alex Kendall, Hayk Martirosyan, Saumitro Dasgupta, Peter Henry, Ryan Kennedy, Abraham Bachrach, and Adam Bry. End-to-end learning of geometry and context for deep stereo regression. *CoRR*, abs/1703.04309, 2017. URL `http://arxiv.org/abs/1703.04309`.

[37] Junhwan Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. volume 2, pages 1033–1040 vol.2, 11 2003. ISBN 0-7695-1950-4. doi: 10.1109/ICCV.2003.1238463.

[38] Teahyung Lee and David Anderson. Performance analysis of a correlation-based optical flow algorithm under noisy environments. pages 4 pp. – 4702, 06 2006. doi: 10.1109/ISCAS.2006.1693679.

[39] Victor Lempitsky, Stefan Roth, and Carsten Rother. Fusionflow: Discrete-continuous optimization for optical flow estimation. In *2008 IEEE Confer-*

ence on Computer Vision and Pattern Recognition, pages 1–8, 2008. doi: 10.1109/CVPR.2008.4587751.

[40] Lahav Lipson, Zachary Teed, and Jia Deng. Raft-stereo: Multilevel recurrent field transforms for stereo matching. *CoRR*, abs/2109.07547, 2021. URL `https://arxiv.org/abs/2109.07547`.

[41] Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. Vancouver, British Columbia, 1981.

[42] Sky McKinley and Megan Levine. Cubic spline interpolation. *College of the Redwoods*, 45(1):1049–1060, 1998.

[43] Moritz Menze, Christian Heipke, and Andreas Geiger. Joint 3d estimation of vehicles and scene flow. In *ISPRS Workshop on Image Sequence Analysis (ISA)*, 2015.

[44] Truong Nguyen Xuan and Huyk-Jae Lee. Energy-based iterative cost aggregation in depth estimation with a stereo camera. pages 319–320, 10 2016. doi: 10.1109/ISOCC.2016.7799806.

[45] Guang-Yu Nie, Ming-Ming Cheng, Yun Liu, Zhengfa Liang, Deng-Ping Fan, Yue Liu, and Yongtian Wang. Multi-level context ultra-aggregation for stereo matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[46] M. Okutomi and Takeo Kanade. A multiple-baseline stereo. *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 63–69, 1991.

[47] J. Anthony Parker, Robert V. Kenyon, and Donald E. Troxel. Comparison of interpolating methods for image resampling. *IEEE Transactions on Medical Imaging*, 2 (1):31–39, 1983. doi: 10.1109/TMI.1983.4307610.

[48] Anurag Ranjan and Michael J. Black. Optical flow estimation using a spatial pyramid network. *CoRR*, abs/1611.00850, 2016. URL `http://arxiv.org/abs/1611.00850`.

[49] Jérôme Revaud, Philippe Weinzaepfel, Zaïd Harchaoui, and Cordelia Schmid. Epicflow: Edge-preserving interpolation of correspondences for optical flow. *CoRR*, abs/1501.02565, 2015. URL `http://arxiv.org/abs/1501.02565`.

[50] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision*, 47(1–3): 7–42, apr 2002. ISSN 0920-5691. doi: 10.1023/A:1014573219977. URL `https://doi.org/10.1023/A:1014573219977`.

[51] Syed Tafseer Haider Shah and Xiang Xuezhi. Traditional and modern strategies for optical flow: an investigation. *SN Applied Sciences*, 3(3):1–14, 2021.

[52] Prasan Shedligeri, Florian Schiffers, Sushobhan Ghosh, Oliver Cossairt, and Kaushik Mitra. Selfvi: Self-supervised light-field video reconstruction from stereo video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2491–2501, October 2021.

[53] Changming Sun. Fast optical flow using cross correlation and shortest-path techniques. 06 2000.

[54] Deqing Sun, Stefan Roth, and Michael Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision*, 106:115–137, 01 2014. doi: 10.1007/s11263-013-0644-x.

[55] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume, 2018.

[56] Zachary Teed and Jia Deng. RAFT: recurrent all-pairs field transforms for optical flow. *CoRR*, abs/2003.12039, 2020. URL `https://arxiv.org/abs/2003.12039`.

[57] Federico Tombari, Stefano Mattoccia, Luigi Di Stefano, and Elisa Addimanda. Classification and evaluation of cost aggregation methods for stereo correspondence. 06 2008. doi: 10.1109/CVPR.2008.4587677.

[58] Guangming Wang, Chi Zhang, Hesheng Wang, Jingchuan Wang, Yong Wang, and Xinlei Wang. Unsupervised learning of depth, optical flow and pose with occlusion from 3d geometry. *CoRR*, abs/2003.00766, 2020. URL `https://arxiv.org/abs/2003.00766`.

[59] Philippe Weinzaepfel, Jérôme Revaud, Zaid Harchaoui, and Cordelia Schmid. DeepFlow: Large displacement optical flow with deep matching. In *ICCV - IEEE International Conference on Computer Vision*, pages 1385–1392, Sydney, Australia, December 2013. IEEE. doi: 10.1109/ICCV.2013.175. URL `https://hal.inria.fr/hal-00873592`.

[60] Haofei Xu and Juyong Zhang. Aanet: Adaptive aggregation network for efficient stereo matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[61] Yao Yao, Zixin Luo, Shiwei Li, Tian Fang, and Long Quan. Mvsnet: Depth inference for unstructured multi-view stereo. *CoRR*, abs/1804.02505, 2018. URL `http://arxiv.org/abs/1804.02505`.

[62] Feihu Zhang and Benjamin W. Wah. Fundamental principles on learning new features for effective dense matching. *IEEE Transactions on Image Processing*, 27 (2):822–836, 2018. doi: 10.1109/TIP.2017.2752370.

[63] Feihu Zhang, Victor Adrian Prisacariu, Ruigang Yang, and Philip H. S. Torr. Ga-net: Guided aggregation net for end-to-end stereo matching. *CoRR*, abs/1904.06587, 2019. URL `http://arxiv.org/abs/1904.06587`.

[64] Feihu Zhang, Xiaojuan Qi, Ruigang Yang, Victor Prisacariu, Benjamin W. Wah, and Philip H. S. Torr. Domain-invariant stereo matching networks. *CoRR*, abs/1911.13287, 2019. URL `http://arxiv.org/abs/1911.13287`.