Some Problems in Depth Enhanced Video Processing

Vanel Lazcano

u*pf.* Universitat Pompeu Fabra *Barcelona*

To My Family.

Agradecimientos

Quisiera agradecer a la profesor Coloma Ballester por toda sus enseñanzas, ayuda, consejos, empuje, tolerancia, paciencia y apoyo todos estos años. Aunque que ya no está con nosotros, quisiera agradecer de igual forma al profesor Vicent Caselles por toda su paciencia, enseanzas, consejos para vivir la vida, haberme integrado a su equipo de trabajo, toda la ayuda para soportar el doctorado, conversaciones sobre política, cine y por su sentido del humor.

Agradecer a Pablo Arias por haber trabajado conmigo un año (gracias por tu paciencia). Quisera agradecer a Luis Garrido por su ayuda, todos sus concejos y sus todos sus observaciones. Agradecer a Felipe Calderero por haber trabajado conmigo y ayudarme con ropa, comida y amistad.

Quisiera agradecer a mis amigos a Enric y Cris, Gabriele y Laura, Pablo y Gloria, Felipe y Annamaria.

Quisera agradecer a los compañeros de despacho y del doctorado: Felipe Calderero, Pablo Arias, Gloria Haro, Margarita Camacho, Enric Meinhardt, Gabriele Facciolo, Rida Sadek, JuanFran Garamendi, Roberto P. Palomares, Edoardo Provenci, Marcelo Bertalmio, Juan Calvo, Vadim, Babak, Christiane Poeschl, María Oliver. A los amigos de la segunda planta: Oriol Martínez, Luis Ferraz. Mi amiga de la tercera planta Verónica Moreno.

Agradezco a los amigos la cuarta planta: Alicia, Simón, Gabriela Ferraro.

Agredecer a los amigos de Barcelona Media: Constantinos Constantinopulos, Aurelie Bugeau, Nicolas Papadakis, Daniel Sciretti, Antonio Baeza, Alun, Irene Sancho, Natalie, Pau Gargallo, Liun Quiangliu, Rosa Zamora,

Abstract

In this thesis we tackle two problems, namely, the data interpolation problem in the context of depth computation both for images and for videos, and the problem of the estimation of the apparent movement of objects in image sequences. The first problem deals with completion of depth data in a region of an image or video where data are missing due to occlusions, unreliable data, damage or lost of data during acquisition. In this thesis we tackle it in two ways. First, we propose a non-local gradient-based energy which is able to complete planes locally. We consider this model as an extension of the bilateral filter to the gradient domain. We have successfully evaluated our model to complete synthetic depth images and also incomplete depth maps provided by a Kinect sensor.

The second approach to tackle the problem is an experimental study of the Biased Absolutely Minimizing Lipschitz Extension (biased AMLE in short) for anisotropic interpolation of depth data to big empty regions without information. The AMLE operator is a cone interpolator, but the biased AMLE is an exponential cone interpolator which makes it more addapted to depth maps of real scenes that usually present soft convex or concave surfaces. Moreover, the biased AMLE operator is able to expand depth data to huge regions. By considering the image domain endowed with an anisotropic metric, the proposed method is able to take into account the underlying geometric information in order not to interpolate across the boundary of objects at different depths. We have proposed a numerical model to compute the solution of the biased AMLE which is based on the eikonal operators. Additionally, we have extended the proposed numerical model to video sequences.

The second problem deals with the motion estimation of the objects in a video sequence. This problem is known as the optical flow computation.

The Optical flow problem is one of the most challenging problems in computer vision. Traditional models to estimate it fail in presence of occlusions and non-uniform illumination. To tackle these problems we proposed a variational model to jointly estimate optical flow and occlusion. Moreover, the proposed model is able to deal with the usual drawback of variational methods in dealing with fast displacements of objects in the scene which are larger than the object itself. The addition of a term that balance gradient and intensities increases the robustness to illumination changes of the proposed model. The inclusions of a supplementary matches given by exhaustive search in specifics locations helps to follow large displacements.

Resumen

En esta tesis se abordan dos problemas: interpolación de datos en el contexto del cálculo de disparidades tanto para imágenes como para video, y el problema de la estimación del movimiento aparente de objetos en una secuencia de imágenes. El primer problema trata de la completación de datos de profundidad en una región de la imagen o video dónde los datos se han perdido debido a oclusiones, datos no confiables, datos dañados o pérdida de datos durante la adquisición. En esta tesis estos problemas se abordan de dos maneras. Primero, se propone una energía basada en gradientes nolocales, energía que puede (localmente) completar planos. Se considera este modelo como una extensión del filtro bilateral al dominio del gradiente. Se ha evaluado en forma exitosa el modelo para completar datos sintéticos y también mapas de profundidad incompletos de un sensor Kinect.

El segundo enfoque, para abordar el problema, es un estudio experimental del biased AMLE (Biased Absolutely Minimizing Lipschitz Extension) para interpolación anisotrópica de datos de profundidad en grandes regiones sin información. El operador AMLE es un interpolador de conos, pero el operador biased AMLE es un interpolador de conos exponenciales lo que lo hace estar más adaptado a mapas de profundidad de escenas reales (las que comúnmente presentan superficies convexas, concavas y suaves). Además, el operador biased AMLE puede expandir datos de profundidad a regiones grandes. Considerando al dominio de la imagen dotado de una métrica anisotrópica, el método propuesto puede tomar en cuenta información geométrica subyacente para no interpolar a través de los límites de los objetos a diferentes profundidades. Se ha propuesto un modelo numérico, basado en el operador eikonal, para calcular la solución del biased AMLE.

Adicionalmente, se ha extendido el modelo numérico a sequencias de video.

El cálculo del flujo óptico es uno de los problemas más desafiantes para la visión por computador. Los modelos tradicionales fallan al estimar el flujo óptico en presencia de oclusiones o iluminación no uniforme. Para abordar este problema se propone un modelo variacional para conjuntamente estimar flujo óptico y oclusiones. Además, el modelo propuesto puede tolerar, una limitación tradicional de los métodos variacionales, desplazamientos rápidos de objetos que son más grandes que el tamaño objeto en la escena. La adición de un término para el balance de gradientes e intensidades aumenta la robustez del modelo propuesto ante cambios de iluminación. La inclusión de correspondencias adicionales (obtenidas usando búsqueda exhaustiva en ubicaciones específicas) ayuda a estimar grandes desplazamientos.

Preface

Motivation

The objective of disparity completion is to complete the lack of information in a region in which data are missing due to problems occurred during acquisition, damage or occlusion. Disparity completion is an important task in many applications in image processing such as construction of elevation models from level curves Almansa et al. (2002), completion of 3D surfaces Digne et al. (2010), super resolutions applications Yang et al. (2007), generation of a virtual view based on reference image and depth information Ndjiki-Nya et al. (2010).

Depth data either acquired by Time-of-Flight cameras or computed using an stereo algorithm, provides a sparse disparity map of reliable data which can be improved by filtering. Inspired by the seminal paper of Perona and Malik (1990) many work proposed to filter a sparse depth map using an anisotropic filter Yin and Cooperstock (2004), Digne et al. (2010). In Digne et al. (2010), the authors propose to diffuse depth values using an iterated bilateral filter with weights given by the similarities of the reference image. The bilateral filter was used also to compute correlation cost Yoon and Kweon (2006) or to obtain a super resolution image Yang et al. (2007) applying iterated bilateral filter.

A depth completion model based on bilateral filter presents one drawback, the incapability to complete large areas of incomplete depth data. An interesting alternative to tackle this drawback is the biased Absolutely Minimizing Lipschitz Extensions filter. AMLE and biased AMLE operator appear as the simplest interpolators that satisfy a set of axioms Caselles et al. (2006).

AMLE filter was discovered originally by Aronsson in Aronsson (1967) and Aronsson (1968) as an interpolation tool from a theoretical point of view.

The actual commercial interest in exhibiting 3D movies has motivated the extension of disparity completion to depth video completion. The main idea is to perform depth completion in the actual frame using depth information contained in previous or posterior frames. The position of depth data information contained in previous or posterior frames should be compensated considering apparent motion of the objects in the video sequence. In general this apparent motion is not known and should be estimated for the video sequence. This motion estimation should be precise enough in such a way that the depth information compensated from other frames will be located in the right position for the posterior depth completion. The apparent motion of the objects in the video sequence is known as the optical flow.

Optical flow estimation is one of the most challenging problem in computer vision, specially in real scenarios where occlusions and illumination changes occur. It has many applications, including autonomous flight of vehicles Ruffier and Franceschini (2005), Kendoul et al. (2005), slow camera motion generation, video deblurring, noise suppression Stiller and Konrad (1999), video compression and many more.

In the seminal work of Horn and Schunck (1981) is presented a model that estimates a dense optical flow based on two assumptions: the brightness constancy assumption and a smooth spatial variation of the optical flow. The optical flow obtained by this model is very smooth and sensitive to the presence of noise. Nevertheless it has inspired many proposals that focus on accuracy Brox et al. (2004a); Black and Anandan (1996); Sun et al. (2010a); Zach et al. (2007), robust terms for brightness constancy assumption L^2 or L^1 or for the smooth spatial variation: isotropic diffusion, image-adaptive, anisotropic diffusion.

In general in real scenarios optical flow estimations methods presents some drawbacks:

- Methods may fail in occlusion areas due to relative motion of the objects in the scene.

- Shadows or light reflexions that appear and move in the image sequence can make the brightness constancy assumption to fail.

- Traditional optical flow methods can not handle displacement of the object in the scene if it is greater than the size of the object itself.

The importance of these two topics in media applications as well as the number of publications appeared in recent years in these topics let us consider to study the problem of disparity completion and robust optical flow. This thesis proposes two methods for depth completion and an illumination changes and large displacement robust optical flow method. It is organized as follows. In Part I we present a non-local gradient-based energy for interpolating incomplete depth maps. It represents an extension of the bilateral filter adapted to reconstruct locally planar disparity maps. In Part II, we performed an experimental study of the biased-AMLE operator for completion of depth maps. We evaluated the performance of the method in upsampling task, completion of depth video sequences and completion of depth maps for large empty regions. In Part III we proposed a variational model for joint optical flow and occlusion estimation, which is adapted for both gray and color image sequences and is able to handle illumination changes as well as large displacements. We close this thesis with conclusions and future work presented in Part IV.

The next sections review in more detail the contributions of each part of the thesis.

Part I: Non-local energy for depth completion

In this part we propose a non-local gradient-based model for interpolating incomplete disparity maps. We apply this model to complete depth data either acquired by Time-of-Flight cameras or computed using an stereo algorithm. The proposed model can be shown to converge to an energy involving second order derivatives, therefore the model presents the ability to obtain higher order interpolations. By including an edge map our model permits also to recover depth discontinuities.

Let us state some notation. We consider a reference image $I(x) : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^3$. We also considered a depth map $u(x) : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ and a confidence map $\beta(x) : \Omega \subset \mathbb{R}^2 \to [0, 1]$. We also consider a edge map $\sigma(x) : \Omega \subset \mathbb{R} \to \mathbb{R}$.

In Chapter 2 we proposed a model for the problem of depth interpolation:

$$E(u) = D(u) + \lambda R_{\mathcal{B}}(u) + (1 - \lambda)R_{\mathcal{P}}(u). \tag{1}$$

Where,

$$D(u) = \int_\Omega \beta(x)\left|u(x) - u^0(x)\right|^2 dx, \tag{2}$$

$$R_\mathcal{B}(u) = \int_\Omega \int_\Omega w_\mathcal{B}(x,y)\left|u(x) - u(y)\right|^2 dx, \tag{3}$$

$$R_\mathcal{P}(u) = \int_\Omega \int_\Omega w_\mathcal{P}(x,y)\|\nabla u(x) - \nabla u(y)\|^2 dx, \tag{4}$$

Equation (2) is the attachment to the initial depth data u^0, equation (3) is a first order non-local regularizer and equation (4) is a second order non-local regularizer. Where λ is a constant value $\in [0,1]$.

The term (3) is a non-local energy enforcing depth self similarities according to the reference image's self-similarity. The weights $w(x,y)$ are computed in terms of the distance and color similarity between x and y are given by:

$$w_\mathcal{B} = \exp\left(-\frac{\|x-y\|}{2k^2}\right)\exp\left(-\frac{|I(x) - I(y)|}{2h^2}\right) \tag{5}$$

The term (4) is a non-local energy that imposes gradient similarity based on color similarity and pixel proximity. The main assumption is that surfaces in the image are locally planar therefore, nearby pixels in the same object have similar depth gradient. To allow depth discontinuities we define the weights $w_\mathcal{P}(x,y) = w_\mathcal{B}\sigma(x)\sigma(y)$, where σ vanishes at strong image edges:

$$\sigma(x) = \frac{1}{1 + e^{((\|\nabla I(x)\| - s_0)/\tau)}} \tag{6}$$

We obtained the Euler-Lagrange equation and it is solved by iteration of a second order Poisson equation.

In Chapter 3 we first present results obtained in a synthetic image. In Figure 1 is shown a synthetic image with two planes. We added Gaussian noise with a standard deviation of 10 to the disparity. We set $\beta = 0$ in the interpolation domain and $\beta = 0.5$ weighting the data attachment. In Figure 2 we compare the profiles of the solutions obtained with bilateral filter ($\lambda = 1$) and gradient-based energies ($\lambda = 0$).

Figure 1: Synthetic depth data (the black region indicates the interpolation domain or interpolation mask), reference image, result using as $\sigma(x)$ the edge map of the reference image given by equation (6), result using bilateral filter.
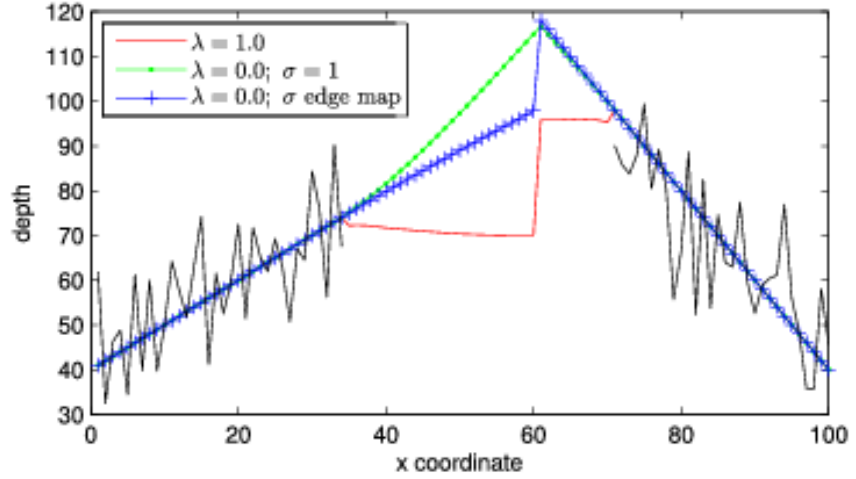

Figure 2: Profile of the result for the two planes synthetic image. We show 4 curves: Noisy data, output of bilateral filter, output of gradient-based with $\sigma = 1$, output of gradient-based with σ given by (6).

We assume that the discontinuities of the disparity map correspond to the discontinuities of the reference image. Therefore, the edge map is computed on the reference image and determines the map $\sigma(x)$. In Figure 2 we show a profile of the interpolated disparity in the given mask (interpolation domain). The model filters the disparity map outside the mask and interpolates it inside. The Figure shows the results of the bilateral filter ($\lambda = 1$, in red), the model considering only gradient ($\lambda = 0$) without the edge map (in green), and the model considering gradient with the edge map (in blue). The Figure shows a profile of the interpolated result. We see that the result for $\lambda = 1$ is a piecewise constant map and does not continue the two planes.

With $\lambda = 0$ and $\sigma = 1$, the result continues the two planes but with a large error. With $\lambda = 0$ and $\sigma(x)$ given by (6), we can continue the slopes of the planes and get the discontinuity. The gradient-based solution considering the discontinuity map performs better than bilateral filter.

Taking into account results obtained in Chapter 4 of the thesis we conclude:

- we exploit the given image both to guide the interpolation and to infer the location of depth discontinuities.

- The model presents the ability to interpolate planes. The proposed model can be regarded as an extension to gradient domain of the bilateral filter.

- The Euler-Lagrange equation can be seen as a non-local fourth order equation, which is solved as an iteration of a second order Poisson equation.

Part II: Biased AMLE operator for depth completion

In this Part we present a numerical scheme and an experimental study of an operator for interpolating incomplete or sparse depth maps. The operator is the biased Absolutely Minimizing Lipschitz Extension that appeared in the axiomatic analysis of the interpolation operators in manifolds in Caselles et al. (2006) and also in the context of tug-of-war games in Peres et al. (2010) (see also Armstrong et al. (2009)).

We are motivated by the extensive literature on filters to enhance sparse depth data (coming from several depth sensors) that has been published in the last years.

When a depth map is obtained either as a result of a stereo algorithm or acquired by a Time-Of-Flight camera or by a Kinect sensor, usually the depth map presents areas without information (holes). This lack of information is due to occlusions between objects in the scene or to areas where the depth data is not reliable or simply missing. Those holes or regions without reliable depth information can have big areas which require appropriate interpolation operators to enhance the depth data. The problem of depth interpolation consists in completing the empty depth data regions in a depth image while guarantying compatibility with the information presents in the depth image.

To this goal, in this part of the thesis we considered the biased AMLE model: $\Delta u_\infty + \beta |\nabla u|_\xi = 0$.

This model has the capability to complete large areas of incomplete data due to the inclusion of the gradient term. This interpolation operator exactly fits the incomplete data at known values and it can interpolate values on isolated points.

Many applications of depth completion technique does exist for example to render new views from an image and its depth image to a different virtual view Ndjiki-Nya et al. (2010) or interpolation of digital elevation models (DEM) Almansa et al. (2002) from contour lines.

In Caselles et al. (2006) a set of axioms was introduce which should be satisfied by any interpolation operator on a surface and derive the associated partial differential equation. Then in Caselles et al. (2006) is deduced that any operator which interpolates continuous data given on a set of curves on a surface can be given as the viscosity solution of an elliptic partial differential equation on a manifold (this is the extension of the corresponding results obtained in the plane in Caselles et al. (1998) to manifolds).

In Chapter 8 we present a numerical scheme to solve the biased AMLE. In Oberman (2005b) and Manfredi et al. (2015) is presented a numerical scheme to compute the infinity Laplacian operator based on the eikonal operator.

Given any two points in the grid x, y, then the geodesic distance d_{xy} is defined by:

$$d_{xy} = \inf\{L_g(\gamma) : \gamma \ is \ a \ curve \ joining \ x \ to \ y \ \}. \tag{7}$$

Given a point x in the grid, let $\mathcal{N}(x)$ be a neighborhood of x.

Following Manfredi et al. (2015) the positive eikonal operator on a graph is defined by

$$\|\nabla u(x)\|_x^+ = \max_{y \in \mathcal{N}(x)} \frac{u(y) - u(x)}{d_{xy}}. \tag{8}$$

The negative eikonal operator on a graph is defined by

$$\|\nabla u(x)\|_x^- = \min_{y \in \mathcal{N}(x)} \frac{u(y) - u(x)}{d_{xy}}. \tag{9}$$

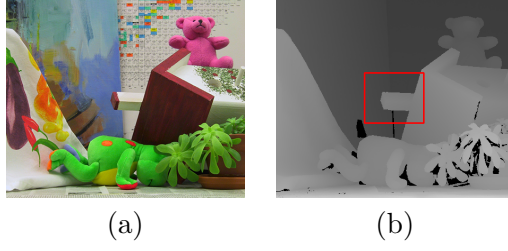<div align="center">(a)   (b)</div>

Figure 3: Original image and depth image for teddy and baby.

The discrete infinity Laplacian corresponds to (Oberman (2005b))

$$\Delta_{\infty,g}u(x) = \frac{\|\nabla u(x)\|_x^+ + \|\nabla u(x)\|_x^-}{2}. \tag{10}$$

Inspired by these numerical schemes we compute a discrete version of the biased AMLE in the form:

$$\frac{\|\nabla u(x)\|_x^+ + \|\nabla u(x)\|_x^-}{2} + \beta\left|\|\nabla u(x)\|_x^+\right| = 0, \tag{11}$$

the solution $u(x)$ depends on the parameter β, $\|\nabla u\|^+$, $\|\nabla u\|^-$ and the sign of the eikonal operator $\|\nabla u(x)\|_x^+$,

$$\frac{\|\nabla u(x)\|_x^+ + \|\nabla u(x)\|_x^-}{2} + \beta\mathrm{sign}(\|\nabla u(x)\|_x^+)\|\nabla u(x)\|_x^+ = 0. \tag{12}$$

Solving for $u(x)$ this numerical scheme, we obtain an iterated version of the biased AMLE:

$$u^{k+1}(x) = \frac{\beta_+ d_{xz}u^k(y) + \beta_- d_{xy}u^k(z)}{\beta_+ d_{xz} + \beta_- d_{xy}}, \qquad k = 0, 1, ... \tag{13}$$

where $\beta_+ = \frac{1}{2} + \beta\mathrm{sign}(\|\nabla u(x)\|_x^+)$ and $\beta_- = \frac{1}{2}$.

In Chapter 9 we present the evaluation of the biased AMLE in the upsampling task. We have taken the Middlebury standard database Baker et al. (2011) that contains a reference color image and its depth image. We show in Figure 3 an example of the database.
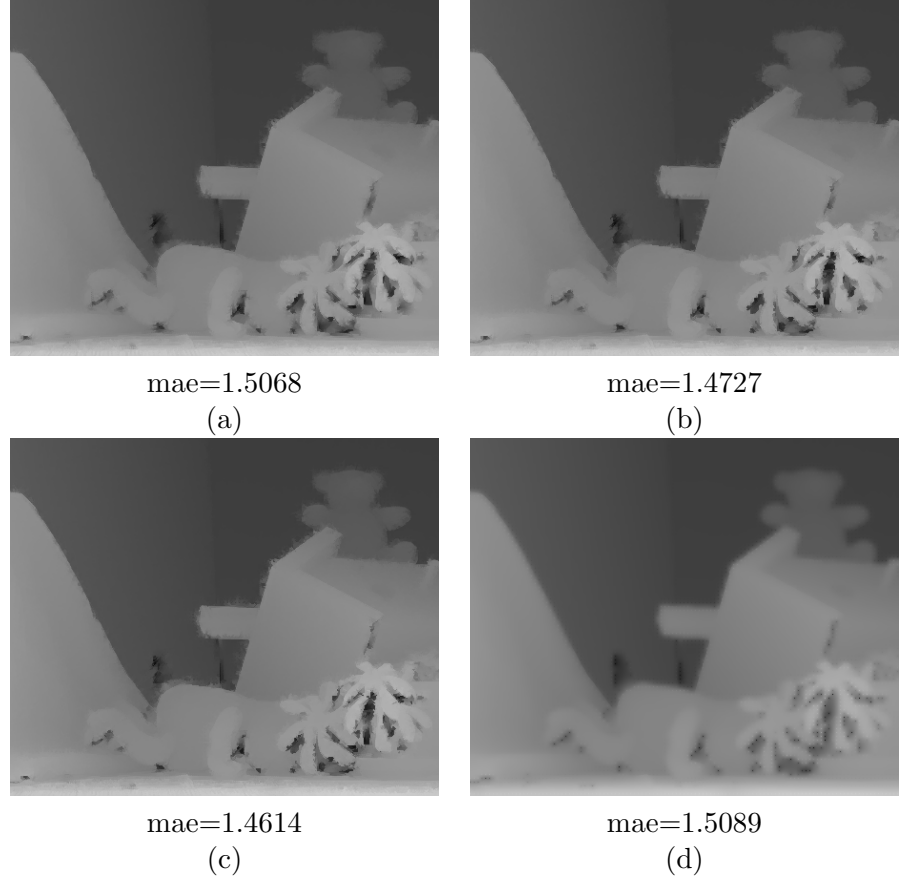
mae=1.5068   mae=1.4727

(a)   (b)

mae=1.4614   mae=1.5089

(c)   (d)

Figure 4: Results of Upsampled task for an image of Middlebury database. (a) Results obtained by AMLE. (b) Biased AMLE $\beta = 1.1$. (c) Biased AMLE $\beta = 1.2$. (d) Bilateral filter.

We took one sample every 16 square pixels. Using this sampled version of the depth image we upsample the data using the AMLE, biased AMLE and bilateral filter as we show in Figure 4.

We show in Figure 4 results in upsampling task obtained by AMLE, biased AMLE, bilateral filter. We have evaluated numerically the performance of each method. We have evaluated the interpolation comparing the upsampled image with the original depth image. We called this error mae (mean of absolute error). We observe that the minimum *mae* value was obtained by biased AMLE $mae = 1.4615$ with $\beta = 1.2$. We observe that using bilateral filter the upsampled depth image is very blur.

In Table 1 we show results numerical of the upsampling task with differente sample rate.

Table 1: Results for AMLE filter, biased AMLE filter and for bilateral filter to subsample teddy depth image

Size	AMLE	biased AMLE $\beta^+ = 1.1$	biased AMLE $\beta^+ = 1.2$	bilateral
Filtered 16×16	2.5749	2.5798	2.5988	2.5535
Filtered 8×8	2.0304	2.0010	2.0019	2.0935
Filtered 4×4	1.5068	1.4727	1.4614	1.5089
Filtered 2×2	1.0884	1.0891	1.0748	1.0288
Average	1.8001	1.7857	**1.7842**	1.7962

Results presented in Table 1 shown that the biased AMLE outperforms the AMLE operator and the bilateral filter in the upsampling task. We have extended the biased AMLE filter to temporal domain. We have tested in a synthetic depth video sequence where we have at our disposal the reference image, the depth map and the optical flow. Given that the sequence is synthetic the depth map and the optical flow do not contains error neither noise. We have added holes to the first depth image of the sequence and we have propagate them to the rest depth image of the sequence, creating a tube in the sequence. We have completed the holes in the video sequence using biased AMLE. We demonstrated that using the optical flow helps to improve results of the completion.

Taking into account the results we conclude:

- The interpolator is computed as a ratio between two terms: a sum of function values weighted by distances and a sum of distances. This simple operation made the operator easy to implement, simple and fast.

- The numerical implementation considers many very strong approximation that experimentally we have shown that their are valid.

- We have evaluated the Biased AMLE in differents experiments: umsampling, temporal extension. Upsampling experiments show that the biased AMLE and AMLE outperform the bilateral filter and biased AMLE outperforms the AMLE.

- We have extended the biased AMLE to temporal domain. We added a new term to the distance metric. This new term considers the available depth information in the next image. The use of the optical flow helps to improve the performance of the depth interpolation in video sequences.

Part III: Illumination changes and large displacement robust optical flow

Optical flow problem concerns with the estimation of the apparent motion between two consecutive images of a video sequence. Most adopted strategies to solve the optical flow problem use a variational approach. In those variational models, the optical flow computation is stated as an energy minimization problem where the energy has, in general, two terms; namely, the data term and the regularization term.

In Part III, we propose a variational model for joint optical flow and occlusion estimation, which is adapted for both gray and color image sequences and is able to handle illumination changes as well as large displacements.

To better explain the different contributions, we have chosen to present it progressively. First, we propose a joint minimization problem to estimate both optical flow and occlusions while preserving discontinuities of the flow. Our data term is based on the brightness constancy constraint for the case of gray sequences and on the color constancy constraint for color sequences. The color constancy assumption or constraint states that the color of the pixels do not change along the displacement of the object Brox et al. (2004b). The proposed energy model

incorporates information that allows to detect occlusions. This information is based on the divergence of the flow and the energy favors the location of occlusions on regions where this divergence is negative. Assuming that occluded pixels are visible in the previous frame, the optical flow on non-occluded pixels is forward estimated whereas is backwards estimated on the occluded ones.

The optical flow constraint presents some drawbacks: often gray or color constancy assumption is violated due to illumination changes, shadows or reflexions, as well as due to occlusions that appear when objects or the camera move. On the other hand, most of the variational models for optical flow computation use the coarse-to-fine strategy to be able to handle large displacement of the objects in images Stoll et al. (2012). This strat-

egy consists in creating a multiscale pyramid, estimating the optical flow at coarser scales and then refining the solution at finer scales Mémin and Pérez (1998); Brox et al. (2004b); Meinhardt-Llopis et al. (2013). However, the coarse-to-fine strategy is unable to handle large displacements of small objects that move differently from their surroundings. In this Part of th thesis we propose and analise a methodology to handle all these drawbacks.

List of Publications

The work in this thesis has led to the publication of the following articles:

- V. Lazcano, P. Arias and V. Caselles. *A gradient based neighborhood filter for depth interpolation.* In *IEEE International Conference in Image Processing(ICIP)*, pages 873–876, 2012.

- P. Arias, V. Caselles, G. Facciolo, V. Lazcano, R. Sadek. *Nonlocal Variational Models for Inpainting and Interpolation.* In *Mathematical Models and Methods in Applied Sciences*, pages 1–64. 2012.

- C. Ballester, L. Garrido, V. Lazcano and V. Caselles, *A tv-l1 Optical Flow Method with Occlusion Detection.* In *DAGM/OAGM, LNCS 7476.* Springer Verlag. 2012.

- J. Garamedi, C. Ballester, L. Garrido, V. Lazcano, *Joint TV-L1 Optical Flow Occlusion Estimation.* In IPOL Journal - Image Processing On Line. In revision, 2015.

- V. Lazcano, C. Ballester, L. Garrido, V. Caselles, *Robust Optical Flow for Large Displacement.* preprint. 2015.

- P. Arias, C. Ballester, F. Calderero, V. Caselles, V. Lazcano. *On geodesic interpolation models.* preprint. 2015.

Contents

PART I

Non-local energy for depth completion

I

ntroduction to the problem

In this part we propose a non-local gradient-based energy for interpolating incomplete disparity maps. It represents an extension of the bilateral filter adapted to reconstruct locally planar disparity maps. Our method is a variational framework for non-local interpolation that exploits the selfsimilarity of natural images to copy information in a consistent way from the known parts of the image. We assume that we have at our disposal a reference image from which similarity weights can be computed.

In the depth interpolation problem, the similarity weights are computed using the reference image and the energy functional imposes that depths should be similar for two nearby pixels that belong to the same object.

When the spatial extend of the weights tends to zero, the proposed model can be shown to converge to an energy involving second order derivatives, explaining thus its ability to obtain higher order interpolations. The proposed energy can be minimized by solving its Euler-Lagrange equation via an iteration of second order Poisson equations. By including an edge map our model permits also to recover depth discontinuities.

1.1 Introduction

In this part we propose an interpolation method for incomplete (or sparse) depth maps based on a gradient domain extension of the bilateral filter (BF) Tomasi and Manduchi (1998). We assume that we are given both an incomplete depth map and a corresponding reference image. The depth

map could be either obtained as a result of a stereo algorithm, or acquired by a camera sensor, like a Kinect sensor or Time of Flight camera (ToF). This depth map may be noisy or unreliable and one needs to filter it or to interpolate a complete one. A ToF camera or a Kinect sensor also give an image, taking it as a left image of the stereo pair, the depth map can be used to generate the right view of the scene for 3D display.

In the task of interpolating disparity maps weights are computed using the reference image. We propose an energy functional which imposes that depths should be similar for two nearby pixels that belongs to the same objects. This is the basic priciple of the bilateral filter Tomasi and Manduchi (1998) which has been adapted to address the depth interpolation problem Yoon and Kweon (2006), Yang et al. (2007), Digne et al. (2010). Additionally the edge information of the image can be incorporated into the depth map, ensuring that discontinuities in depth are consistent with gray level (or color) discontinuities.

We remark that BF computes at a pixel x a weighted average of the given depth map at pixels y on a neighborhood of x, say $\mathcal{N}_x$ (the weights being given in terms of the distance and the color similarity between x and y). In our case we took inspiration by the Poisson model of inpainting Arias et al. (2009, 2011) and we extended it to copy not the disparity but the gradient, enabling the interpolation of planes.

Thus, our model is formulated in terms of a nonlocal energy that compares gradients (or a combination of gradients and intensities) at points where the reference image is similar. In the present formulation, with weights depending only on a reference image, we prove the Γ-convergence of the proposed energy, which is non-local, as we localize the neighborhood where pixels are compared, obtaining the energy

$$\int_\Omega \mathrm{Trace}(D^2u(x)Q(\nabla I(x))D^2u(x))\,dx, \qquad (1.1)$$

where D^2u denotes the Hessian of u and $Q(\nabla I(x))$ is a tensor inhibiting the interpolation across edges (discontinuities) of the image I. This shows that the non-local Poisson energy for disparity interpolation is a non-local version of a second order functional. Moreover, thanks to the anisotropic tensor $Q(\nabla I(x))$ we are able to interpolate planes locally. The interphases separating two different planes are given by the edges of the reference image. Notice that the previous result is an extension of the asymptotic expansion Buades et al. (2006) that identifies the anisotropic diffusion equation underlying the bilateral filter. The proof is based on the results of Bourgain

et al. (2001) approximating Sobolev norms by their non-local version. The application of this energy model to the interpolation of disparity maps acquired using a Kinect sensor together with the generation of the right image of the stereo pair is the object of research in Lazcano et al. (2012).

1.1.1 Related work

Many correlation based stereo matching algorithms provide a sparse disparity map of reliable data which can be improved by filtering. In Yin and Cooperstock (2004), inspired by Perona-Malik equation, the authors propose to use an anisotropic filter that inhibits diffusion accross edges of the reference image. In Digne et al. (2010), the authors propose to diffuse depth values using an iterated BF with weights given by the similarities of the reference image.

The BF has also been used to post-process the similarity (or correlation) cost in stereo matching. In Yoon and Kweon (2006) this similarity measure is computed by bilateral filtering the pixel-based raw matching cost based on color absolute differences, considering BF weights depending on color similarity and geometric proximity to the reference pixel. In this way, one gets an adaptive support window and weights that take into account image discontinuities. In Yang et al. (2007), the authors have used a registered high-quality reference image and the BF to enhance the resolution of low quality depth maps given by range images. In an iterative way and starting from the given range map, an aggregated cost volume is computed by bilateral filtering each slice of an initial cost volume determined by the current depth map.

The non-local Poisson model for image inpainting can be considered as a higher order model (involving a non-local derivative of the gradient). As we have seen it, it permits to create smooth transitions between given data. Inspired by this, we study in this part of the thesis a gradient-domain extension of the bilateral filter Tomasi and Manduchi (1998) (named as neighborhood filter in Yaroslavsky (1985)) for disparity interpolation from incomplete or sparse data for which higher order interpolation may be required. For instance, we may need to interpolate disparity in a given region as a step for stereo inpainting. We may also be given an incomplete disparity map that may have been obtained either as a result of a stereo algorithm or acquired by a camera sensor, like a Time-of-Flight camera (ToF) or Kinect sensor, and a corresponding reference image.

Proposed model and discretization

In this chapter we present our proposed model to solve the problem of disparity completion. As in the bilateral filter, the reference image constraints the comparison of disparity gradients (and/or values) to neighboring pixels that have a similar color. This is encoded by a weight function $w(x, y)$ that is given beforehand and is based on pixel (or patch) comparison. By means of $w(x, y)$ the edge information of the image is incorporated into the disparity map, ensuring that discontinuities in disparity are consistent with gray level (or color) discontinuities. But in contrast to the bilateral filter, that would compute at a pixel x a weighted average of the given disparity map at pixels y on a neighborhood of x, say $\mathcal{N}_x$, (the weights being given in terms of the distance and the color similarity between x and y), we solve a Poisson equation that tries to copy at x the weighted average of gradients at pixels $y \in \mathcal{N}_x$. We will show that use of gradient information permits to extend the planes existing in the given disparity data.

2.1 The Proposed Model

Let us introduce some notation. We consider a reference image $I : \Omega \to \mathbb{R}^n$, where $n = 1$ for a gray level image or $n = 3$ for a color image. As usual, the image domain Ω is a rectangle in $\mathbb{R}^2$. We will also consider a depth map $u : \Omega \to \mathbb{R}$ and a corresponding confidence mask $\beta : \Omega \to [0, 1]$. At each location $x \in \Omega$, $\beta(x)$ measures the confidence we have on the depth value $u(x)$. If we have no data at x, $\beta(x) = 0$.

2.1.1 The iterated bilateral filter

The bilateral filter Tomasi and Manduchi (1998) is a example of neighborhood filter, Yaroslavsky (1985), which takes into account the gray level values of the image to define neighboring pixels. Then the bilateral filter is given by:

$$B(I)(x) = \frac{1}{c(x)} \int_\Omega w_\mathcal{B}(x,y) I(y) \mathrm{d}y, \qquad (2.1)$$

where $w_\mathcal{B} : \Omega \times \Omega \to \mathbb{R}^+$ is given by

$$w_\mathcal{B}(x,y) = \exp\left(-\frac{|x-y|^2}{2k^2} - \frac{\|I(x) - I(y)\|^2}{2h^2}\right), \qquad (2.2)$$

and $c(x) = \int_\Omega w_\mathcal{B}(x,y) \mathrm{d}y$ is the normalization factor, $k, h > 0$. Notice that $w_\mathcal{B}(x,y)$ incorporates both the Euclidean distance of y to the reference pixel x and the *photometric* distance based on the comparison of $I(x)$ and $I(y)$. The spatial extend (euclidean distance) of the neighborhood is controlled by k, whereas the *photometric* extend is controlled by h. Thus the effective extend of the *photometric* neighborhood is given by a packet of level lines $\{y \in \Omega : I(x) - 3h \leq u(y) \leq I(x) + 3h\}$. When h and k are of the same order, the behavior is similar to the Perona-Malik equation Perona and Malik (1990) which inhibits diffusion along large gradients of the reference image Buades et al. (2006).

The bilateral filter has found many applications to different problems in image and surface processing. In particular, to the processing of disparity maps. Indeed, many correlation based stereo matching algorithms provide a sparse disparity map of reliable data which can be improved by filtering. In Yoon and Kweon (2006); Yang et al. (2007); Digne et al. (2010) the authors perform, respectively, disparity super-resolution and interpolation by iterating a bilateral filter on the disparity map u:

$$u^{k+1}(x) = \frac{1}{c(x)} \int_\Omega w_\mathcal{B}(x,y) u^k(y) \mathrm{d}y, \qquad (2.3)$$

where the weights are given by (2.2) and incorporate a registered high-quality reference image I. Thus, disparity values are diffused with weights given by the similarities of the reference image in a neighborhood of x.

The iterative bilateral filtering can be interpreted as a process to minimize the non-local energy functional

$$R_\mathcal{B}(u) = \int_\Omega \int_\Omega w_\mathcal{B}(x,y)(u(x) - u(y))^2 \mathrm{d}y \mathrm{d}x. \qquad (2.4)$$

Energy (2.4) favors constant or piecewise constant minima (on regions separated by high image gradients). This motivates the use of $R_{\mathcal{B}}$ as a regularization term, to be used together with a data attachment term to the initial disparity:

$$E(u) = D(u) + \lambda R_{\mathcal{B}}(u), \qquad \lambda > 0. \tag{2.5}$$

In the context of data interpolation we consider the following data attachment

$$D(u) = \int_{\Omega} \beta(x)|u(x) - u^0(x)|^p \mathrm{d}x, \tag{2.6}$$

where u^0 is a given data, $\beta(x)$ is a confidence mask (or the mask where the data is known), and $p = 1, 2$.

The regularization term $R_{\mathcal{B}}$ is based on the assumption that nearby locations with similar colors should correspond to the same object and have similar disparity since the object's surface is assumed to vary smoothly. Although this assumption works well in many cases, it favors piecewise constant disparity maps (see Figure 3.2).

2.1.2   A gradient based model

Taking account the limitations of the bilateral filter we propose a higher order model so that nearby locations with similar colors have similar *depth gradient*. This is equivalent to assume that objects are locally planar. To achieve this, the corresponding regularization term is

$$R_{\mathcal{P}}(u) = \int_{\Omega} \int_{\Omega} w_{\mathcal{P}}(x,y) \|\nabla u(x) - \nabla u(y)\|^2 \mathrm{d}y \mathrm{d}x.$$

Here $w_{\mathcal{P}}(x,y)$ denotes a slight modification of the BF weights $w_{\mathcal{B}}$ with the aim of obtaining depth maps with discontinuities (or with high gradients).

In practice it is convenient to combine both regularization terms. Thus, we define the energy

$$E(u) = D(u) + \lambda R_{\mathcal{B}}(u) + (1 - \lambda) R_{\mathcal{P}}(u), \quad \lambda \in [0, 1].$$

2.1.3   The discontinuity map

Without a proper choice of the weights, the gradient-based regularization term, $R_{\mathcal{P}}$ would not allow to recover discontinuities in the depth map (in the continuous setting, a discontinuity has infinite energy). By choosing the

weights we remove from $R_{\mathcal{P}}$ the regions of the image where a discontinuity of the depth map is likely to occur. We obtain this information from the edges of the reference image. For that, we define a discontinuity map $\sigma : \Omega \to [0, 1]$ by

$$\sigma(x) = \frac{1}{1 + \exp\left((\|\nabla \hat{I}(x)\|^2 - s_0)/\tau\right)}, \qquad (2.7)$$

where $\hat{I}$ is the reference image smoothed by a Gaussian kernel. This is a (decreasing) soft-thresholding function around s_0 where $-(2\tau)^{-1}$ is the slope of the function at s_0. We incorporate the σ into the weights, defining a new weight function:

$$w_{\mathcal{P}}(x, y) = w_{\mathcal{B}}(x, y)\sigma(x)\sigma(y).$$

Strictly speaking, discontinuities are only allowed when $\sigma(x) = 0$. Still, small non-zero values favor high gradients of the depth map (if needed).

The modified weights $w_{\mathcal{P}}$ are only used in the gradient-based regularization term $R_{\mathcal{P}}$. For $R_{\mathcal{B}}$ we use the original weights $w_{\mathcal{B}}$.

2.1.4   Further justification

Since $u(x) - u(y)$ can be considered as a nonlocal gradient and $\nabla u(x) - \nabla u(y)$ as a non-local second order operator, by suitably rescaling the energies we can localize them and reveal what are the underlying differential operators. To this aim, let us introduce a scale factor $\epsilon > 0$ and define $\rho_S^\epsilon(x) = \frac{1}{\epsilon^2}\rho_S(\frac{x}{\epsilon})$, and $w_{\mathcal{B}}^\epsilon(x, y) = \rho_S^\epsilon(x - y)\rho_I(\frac{I(x) - I(y)}{\epsilon})$.

We also consider rescaled versions of the energies:

$$R_{\mathcal{B}}^\epsilon(u) = \frac{1}{\epsilon^2} \int_\Omega \int_\Omega w_{\mathcal{B}}^\epsilon(x, y)(u(x) - u(y))^2 \mathrm{d}y\mathrm{d}x,$$

$$R_{\mathcal{P}}^\epsilon(u) = \frac{1}{\epsilon^2} \int_\Omega \int_\Omega w_{\mathcal{P}}^\epsilon(x, y)\|\nabla u(x) - \nabla u(y)\|^2 \mathrm{d}y\mathrm{d}x.$$

Let us introduce some notation, necessary to state the result. For a given $v \in \mathbb{R}^2$ we define the 2×2 matrix $Q(v)$ as: $Q(v) = \int_{\mathbb{R}^2} \rho_S(z)\rho_I(\langle v, z \rangle)z \otimes z \, dz$, where $z \otimes z = zz^T = (z_i z_j)_{i,j=1}^2$. Note that $Q(v)$ is defined using the non-rescaled similarity kernels ρ_I and ρ_S. By explicit computation, we have

$$Q(v) = 2\pi k^4 \left(1 + \frac{k^2}{h^2}|v|^2\right)^{-1/2} \left(I + \frac{k^2}{h^2} v \otimes v\right)^{-1}.$$

The following Theorem expresses the convergence of the 'non-local' energies $R_{\mathcal{B}}^{\epsilon}(u), R_{\mathcal{P}}^{\epsilon}(u)$ to a corresponding 'local' version. This permits to show what is the underlying differential operator (hence, the order of interpolation). The corresponding result holds if we add the data terms.

Theorem 2.1. *As $\epsilon \to 0+$, the energies $R_{\mathcal{B}}^{\epsilon}$ Γ-converge to the energy*

$$R_{\mathcal{B}}^0(u) := \int_{\widetilde{\Omega}} \nabla u(x)^T Q(\nabla I(x)) \nabla u(x) \, dx,$$

As $\epsilon \to 0+$, the energies $R_{\mathcal{P}}^{\epsilon}$ Γ-converge to the energy

$$R_{\mathcal{P}}^0(u) := \int_{\widetilde{\Omega}} \mathrm{Tr}(D^2 u(x) Q(\nabla I(x)) D^2 u(x)) \, dx,$$

where $\mathrm{Tr}(A)$ denotes the trace of the matrix A.

We will not give the precise definition of Γ-convergence, it suffices to say that with this notion the minima of the approximating energies converge to minima of the limit one Braides (2002). We have denoted by $D^2 u(x)$ the Hessian matrix of u at x. We give the proof of Theorem 2.1 in the appendix A.0.1.

Notice that $Q(\nabla I)$ is small for large values of $|\nabla I(x)|$. Assuming that it is zero at a set of curves S given by the edges of the image, the energy of piecewise affine functions u with eventual jumps at S is zero. This explains the ability of the proposed model to obtain piecewise affine reconstructions.

2.2   Discretization

We discretize the energy using a finite differences scheme. We consider a rectangular lattice $\Omega_d \subset \mathbb{Z}^2$. For a discrete depth image $u : \Omega_d \to \mathbb{R}$, we define the discrete gradient as a forward difference operator:

$$[\nabla^+ u(x)]_i = \begin{cases} u(x + e_i) - u(x) & \text{if } x + e_i \in \Omega_d \\ 0 & \text{else,} \end{cases}$$

with $i = 1, 2$, where $e_1 = (1, 0)$ and $e_2 = (0, 1)$ and $[\nabla^+ u]_i$ denotes the derivative in the direction e_i. The gradients at the last row and column of the image (set to zero) are removed from the energy by defining $\sigma(x) = 0$ whenever $x + e_i \notin \Omega_d$ for $i = 1$ or $i = 2$.

The discrete gradient-based regularization term results,

$$R_{\mathcal{P},d}(u) := \sum_{x \in \Omega_d} \sum_{y \in \Omega_d} w_{\mathcal{P}}(x,y) \| \nabla^+ u(x) - \nabla^+ u(y) \|^2. \qquad (2.8)$$

Similarly we discretize the other energy terms $R_{\mathcal{B},d}$ and finally obtaining $E_d(u)$ as

$$
\begin{aligned}
E_d(u) = &\sum_{x \in \Omega} \beta(x)(u(x) - u(y)) + \\
&\lambda \sum_{x \in \Omega} \sum_{y \in \Omega} w_{\mathcal{B}}(x,y)(u(x) - u(y))^2 + \\
&(1 - \lambda) \sum_{x in \Omega} \sum_{x \in \Omega} \tilde{w}_{\mathcal{P}}(x,y)(\nabla u(x) - \nabla u(y))^2,
\end{aligned}
$$

The resulting energy $E_d(u)$is quadratic. By computing the Euler-Lagrange equations we obtain a linear system whose corresponding matrix is relatively dense, due to the non-locality of the similarity weights. We solve it using the following fixed point iterations

$$\lambda c(x)u^{k+1} - (1 - \lambda)\mathrm{div}^-(a(x)\nabla^+ u^{k+1}) = f(u^k, x),$$

in which we compute the right hand side using the previous iterate u^k. As initialization we use the given depth data u^0. The new iterate u^{k+1} is computed as the solution of a discrete Poisson equation. We use the conjugate gradient to solve this pentadiagonal system.

The divergence operator div^- is computed as the dual of ∇^+ (i.e. using backward differences), and

$$
\begin{aligned}
c(x) &= \beta(x) + 2\lambda \sum_{y \in \Omega_d} w_{\mathcal{B}}(x,y), \\
a(x) &= 2 \sum_{y \in \Omega_d} w_{\mathcal{P}}(x,y), \\
f(u,x) &= (1 - \lambda)\mathrm{div}^- \mathbf{v}(u,x) + \beta(x)u^0(x) \\
&\quad - 2\lambda \sum_{y \in \Omega_d} w_{\mathcal{B}}(x,y)u(y), \\
\mathbf{v}(u,x) &= \sum_{y \in \Omega_d} w_{\mathcal{P}}(x,y)\nabla^+ u(x).
\end{aligned}
$$

Results

3.1 Results

In this chapter we present we present results for four comparative experiments. The first and second experiment were performed using synthetics data. The third experiment was performed using data form Middlebury stereo database. The fourth experiment was performed using real data obtained by a Kinect sensor for Microsoft X-Box.

3.1.1 Synthetic data

To test our algorithm, we have created a synthetic image with two planes displayed in Figure 3.1. We added Gaussian noise with a standard deviation of 10 to the disparity. We set $\beta = 0$ in the editing domain and $\beta = 0.5$ weighting the data attachment. In Figure 3.2 we compare the profiles of the solutions obtained with BF ($\lambda = 1$) and gradient-based energies ($\lambda = 0$). Note that BF has been used in the literature (Digne et al. (2010), Yang et al. (2007)) to address the problem of depth interpolation/superresolution.

The disparity map (the data) is only known outside the mask and we know the reference image. We assume that the discontinuities of the disparity map correspond to the discontinuities of the reference image. Thus the edge map is computed on the reference image and determines the map $\sigma(x)$. In Figure 3.2 we show a profile of the interpolated disparity in the given mask. We have interpolated the disparity data with a Gaussian noise of standar deviation 5. The model filters the disparity map outside the mask and interpolates it inside. The Figure shows the results of the bilateral
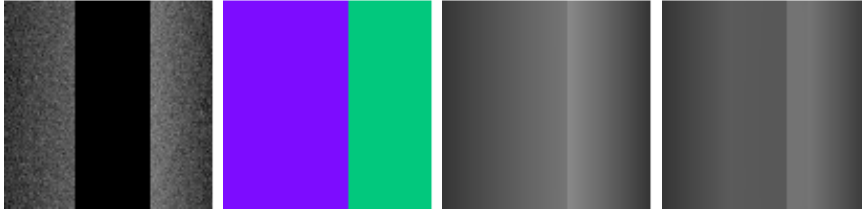
13

Figure 3.1: Synthetic depth data (the black region indicates the interpolation domain), reference image, our result using as $\sigma(x)$ the edge map of the reference image given by (2.7), result using BF.
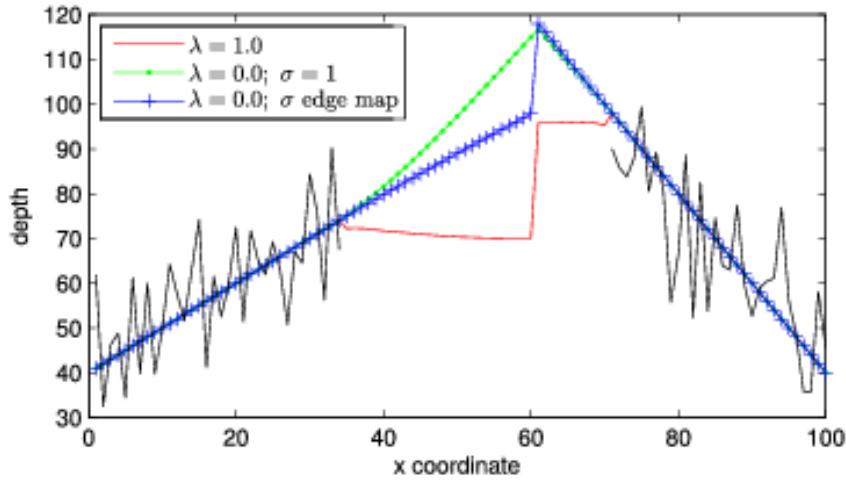

Figure 3.2: Profile of the result for the two planes synthetic image. We show 4 curves: Noisy data, output of BF, output of gradient-based with $\sigma = 1$, output of gradient-based with σ given by (2.7).

filter ($\lambda = 1$, in red), the gradient model ($\lambda = 0$) without the edge map (in green), and the gradient model with the edge map (in blue).

We see that the result for $\lambda = 1$ is a piecewise constant map and does not continue the two planes. With $\lambda = 0$ and $\sigma = 1$, the result continues the two planes but with a large error. With $\lambda = 0$ and $\sigma(x)$ given by (2.7), we can continue the slopes of the planes and get the discontinuity. The gradient-based solution considering the discontinuity map performs better than BF.

A synthetic example of a sphere   In Figure 3.3 we display a synthetic image with a sphere. The data is only known outside the mask (middle of the first row) and we know the reference image from which we have extracted the edge map (right of the first row). We used (2.7) to interpolate the data in the given mask. In the second row of Figure 3.3 we show a profile of the results: we display the result obtained using the gradient model ($\lambda = 0$) and the result obtained using the bilateral filter ($\lambda = 1$). In both cases we display the profile of the original sphere and the interpolated result.
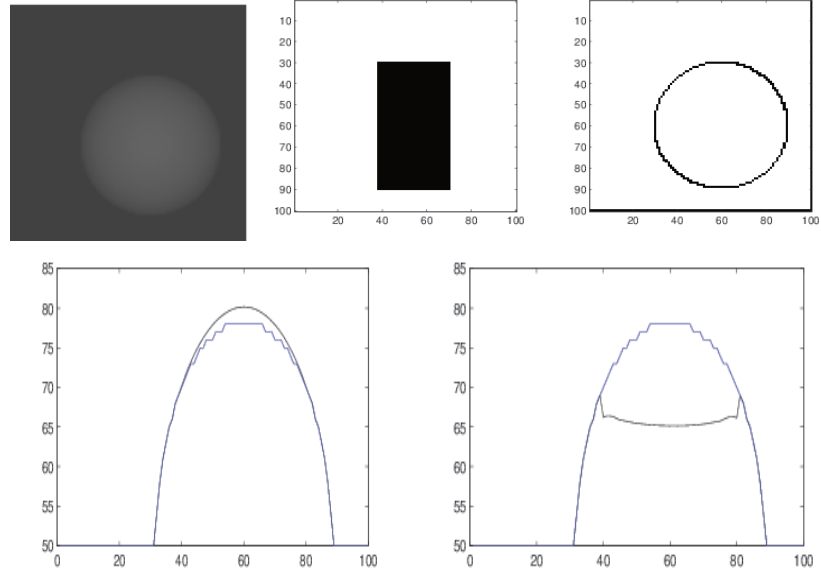

Figure 3.3: Interpolation of a section of a sphere.

3.1.2   Examples with real data

We present two examples with real test images. In our first experiment, we are given a disparity map with a hole corresponding to an object that has been eliminated and our purpose is to interpolate its disparity. The data has been taken from the Middlebury database Scharstein and Szeliski (2002) and is shown in Figure 3.4.

Note that the reference image is unknown in the interpolation domain. To create the discontinuity map we extend the significant level lines Desolneux et al. (2001) arriving at the hole, using Euler's spiral as in Cao et al. (2011). The extended lines determine five regions corresponding to the objects sur-
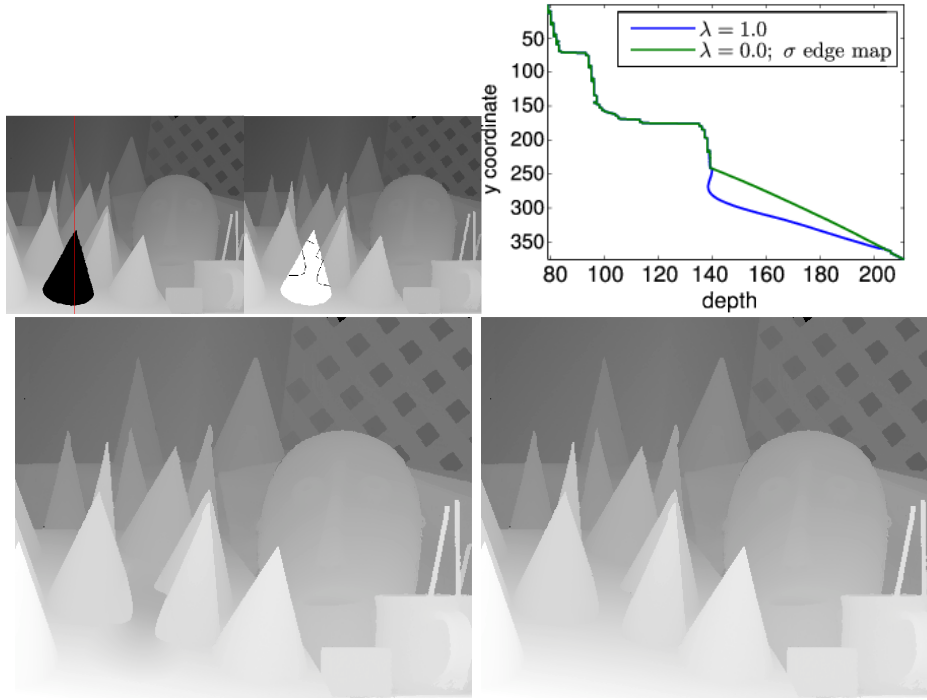
Figure 3.4: Inpainting disparity data for cones image.

rounding the interpolation domain (see Figure 3.4). The similarity weights are set to $w_{\mathcal{B}}(x,y) = 1$ if x and y are in the same region, or 0 otherwise. The results corresponding to BF and gradient-based energies, together with their profiles, are shown in Figure 3.4. The profiles are taken from the column $x = 130$ (indicated as a red line), depicted in the figure ($y = 0$ corresponds to the top of the image). Note how the gradient-based method captures the plane of the table, from $y = 241$ to the end.

In our second experiment with real data we are given a disparity map obtained from Kinect's depth data. The depth map is registered to the reference image using the matlab toolbox provided by Herrera et al. (2011). We take the inverse value of the depth map to obtain the disparity, where data is available. To demonstrate the performance of our model, we interpolate the remaining holes and additionally we suppress an object. Figure 3.5 shows our results obtained with $\lambda = 0.01$, and an edge map $\sigma(x)$ given by Eq. (2.7).
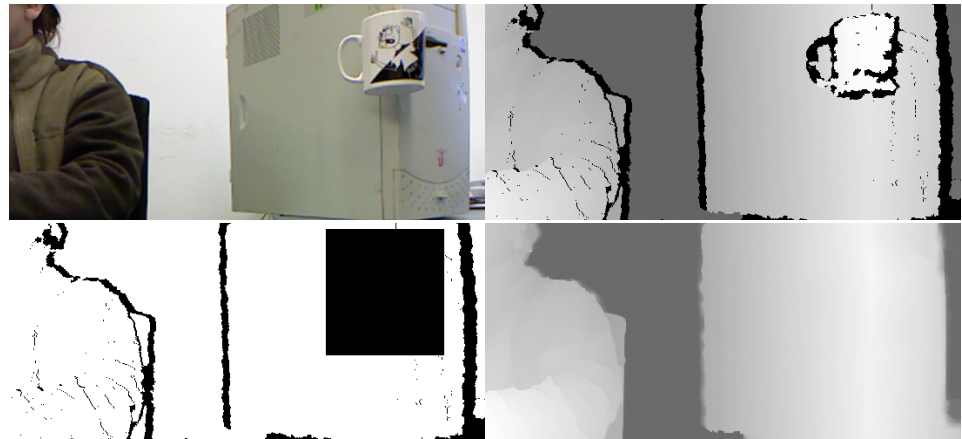
Figure 3.5: Inpainting Kinect's disparity data. Reference image, initial disparity map, mask of interpolation domain, results of the disparity completion using $\lambda = 0.01$, $h = 6$, $k = 5$.

Conclusions

We have presented a non-local gradient-based energy for disparity map interpolation, knowing a reference image. We have taken advantage of the given image both to guide the interpolation and to specify the location of depth discontinuities. We have shown with synthetic examples and two real examples the ability of our method to interpolate planes, concluding that it improves on the bilateral filter. Its Euler-Lagrange equation can be seen as a non-local fourth order equation, which can be solved as an iteration of second order Poisson equations.

The proposed functional can be used as a regularizer in a full stereo algorithm.

Our method presents some limitation in the case of textured images. The completion of the disparity will follow the texture of the reference image. Some regions of the completed disparity will stay isolated of the completion process. Another limitation appears in the case of completion of huge regions of incomplete disparity information. The algorithm converges before the difussion process covers the entire huge regions. This second limitation will be overcome by the method explained in the next part of this thesis.

PART II

Biased AMLE operator for depth completion

Preliminary

In this second part of the manuscript we perform an in-depth experimental study of an operator for interpolating incomplete disparity maps. The operator is the biased Absolute Minimizing Lipschitz Extension that appeared in the axiomatic approach proposed in Caselles et al. (1998) and Caselles et al. (2006). This operator is the simplest interpolation operator satisfying a well-founded set of axioms.

5.1   Introduction

In this work we present a numerical scheme and an experimental study of an operator for interpolating incomplete or sparse depth maps. The operator is the biased Absolute Minimizing Lipschitz Extension that appeared in the axiomatic analysis of the interpolation operators in manifolds in Caselles et al. (2006) and also in the context of tug-of-war games in Peres et al. (2010) (see also Armstrong et al. (2009)).

We are motivated by the extensive literature on filters to enhance sparse depth data (coming from several depth sensors) that has been published in recent years.

When a depth map is obtained either as a result of a stereo algorithm or acquired by a Time-Of-Flight camera or by a Kinect sensor, usually the depth map presents areas without information (ofen referred to as holes). This lack of information may be due to occlusions between objects in the scene or to areas where the depth data is not reliable or simply missing. Those holes or regions without reliable depth information can have big ar-

23

eas which require appropriate interpolation operators to enhance the depth data. The problem of depth interpolation consists in completing the empty depth data regions in a depth image while guarantying compatibility with the information presents in the depth image.

To this goal, in this work we consider the biased AMLE model: $\Delta u_\infty + \beta |\nabla u_\xi| = 0$. This model has the capability to complete large areas of incomplete data due to the inclusion of the gradient term. This interpolation operator exactly fits the incomplete data at known values and it can interpolate values on isolated points.

Depth completion is an active area of interest mainly due to its many applications. For instance, depth completion techniques do exist to render new views from an image and its depth image to a different virtual view Ndjiki-Nya et al. (2010), or for interpolation purposes of digital elevation models (DEM) from contour lines Almansa et al. (2002).

5.1.1 Previous work

The Absolute Minimizing Lipschitz Extension (AMLE) operator was first introduced in Aronsson (1967) and Aronsson (1968) as an interpolation tool from a theoretical point of view. In Caselles et al. (1998) and Caselles et al. (2006), AMLE operator emerges as the simplest operator satisfying a set of axioms.

The uniqueness of viscosity solutions of the AMLE presented in Aronsson (1967, 1968) was proved in Jensen (1993). For a survey we refer to Aronsson et al. (2004); Crandall (2008). The AMLE with variable coefficients was the object of study in Juutinen (1998) and more recently in Gómez and Rossi (2013).

As it is well known Caselles et al. (1998); Crandall et al. (2001); Aronsson et al. (2004), the AMLE is a cone interpolator and is able to interpolate the value at points. The biased AMLE offers an interesting alternative since the value at the points will expand and produce a smoother profile. This is of interest in applications to depth interpolation.

In Almansa et al. (2002) the AMLE is used to interpolate data between level lines. They used it to create elevation reconstruction models from a scanned map where only a few iso-level curves are available.

In Caselles et al. (2006) the AMLE operator was applied to the completion on surfaces embedded in $\mathbb{R}^3$ given the values on a set curves included in the

surfaces.

In Belyaev and Seidel (2005) and in Galic et al. (2015) a new method for image compression is proposed. With this goal they remove the less significant pixels and the remaining pixels serve as scattered interpolation data. The interpolation is performed by a diffusion process based on partial differential equations.

The biased Absolutely Minimizing Lipschitz Extensions appeared in the axiomatic analysis of the interpolation operators in manifolds in Caselles et al. (2006) although was not experimentally studied there (and was also unnoticed in Caselles et al. (1998)). It was also introduced in the context of tug-of-war games in Peres et al. (2010) where it was mathematically studied. It has been also been studied in Armstrong et al. (2009).

AMLE and biased AMLE operators are particularly efficient in order to interpolate sparse disparity maps. This fact is very useful in the case of our interest where we may have huge size disparity holes in a huge size disparity map with few data, or few data with high confidence level.

5.1.2 Summary of contributions

The problem of depth interpolation consists in completing the empty depth data regions in a depth image while guarantying the compatibility with the information presents in the depth image. The biased AMLE operator exactly fits the available data and it can interpolate values on isolated points. This is not satisfied for other interpolation operators like, e.g., the Laplacian. On the other hand, while the AMLE operator is known to be a cone interpolator which is able to interpolate the value at points, the biased AMLE offers an interesting alternative since the value at the points will expand and produce a smoother profile. Thus, the biased AMLE properties make this operator capable to extend the known data to huge empty regions which is of interest in applications to depth interpolation.

In this work, the task of interpolating sparse disparity maps is performed using this reference image and the incomplete disparity map. Indeed, we assume that we have at our disposal a reference image from which an anisotropic metric g and a geodesic distance between points is computed. We propose an algorithm and experimentally study the following equation

$$\Delta_{\infty,g} u + \beta \left| \nabla u \right|_\xi = 0 \quad \text{in } \Omega, \tag{5.1}$$

coupled with the boundary condition

$$u|_{\partial\Omega} = \theta, \tag{5.2}$$

where Ω is a domain in $\mathbb{R}^2$, g is a metric defined on this set which takes into account geodesic spatial distances and photometric similarities, and $\beta \in \mathbb{R}^+$. θ is a continuous function on $\partial\Omega$ providing the known values on $\partial\Omega$.

In order to solve equation 5.1, we propose to use the discretization method proposed in Oberman (2005b) and Manfredi et al. (2015).

We have tackled the problem of anisotropic interpolation of depth data to big empty regions without information. By considering the the image domain endowed with an anisotropic metric, the proposed method is able to take into account the underlying geometric information in order not to interpolate across the boundary of objects at different depths.

On the other hand, we extend the biased AMLE operator to time domain. To do so, we consider the depth data in a video sequence. We experimentally prove that this extension makes the operator capable for completion of depths in depth video sequences.

5.1.3 Structure of Part II

Chapter 6 reviews a set of suitable axioms (already presented in Caselles et al. (2006)) that a simple and efficient interpolator should satisfy.

In Chapter 7 we review the Biased AMLE and we illustrate its behavior on a toy example as a proof of concept .

In Chapter 8 we present a numerical scheme to solve the biased AMLE which is based on the eikonal operator defined in Manfredi et al. (2015) and Oberman (2005b).

Chapter 9 presents the evaluation of the biased AMLE in an upsampling task. Also, we present its extension to temporal video sequences. On the other hand, we provide a comparison of the performance of the biased AMLE with AMLE and with bilateral filter.

In Chapter 10 we compare the performance of the biased AMLE with the gradient-based non-local bilateral presented in the part II of this manuscript.

Finally in Chapter 11 we discuss the obtained results and present our conclussions.

Axiomatic for depth interpolation operators on Manifols

In this chapter we present a set of basics axioms to be satisfied by an interpolation algorithm. Each interpolation algorithm appears as a viscosity solution of a degenerated elliptic partial differential equation. These interpolation methods can be applied in several context: in Digital Elevation Models models from a sample of its level curves, for interpolation of climate maps, and for disparity interpolation. We then focus on one of those interpolation operators: the biased AMLE. Finally we present its extension to temporal video sequences of incomplete depth data.

6.1 Axioms for biased Absolut Minimizing Lipschitz Extension

To complete sparse depth maps, an efficient depth interpolation operator is needed. In this section we consider the general problem of obtaining interpolation operators satisfying a set of suitable axioms and properly defined on Riemannian manifolds. Such manifolds arise for instance for images defined on $\mathbb{R}^2$, endowed with a suitable metric. The authors of Caselles et al. (1998, 2006) studied models and algorithms for interpolating data which is given on a set of curves. Following an axiomatic approach, they obtained that the interpolation operators satisfying a set of suitable axioms are given

in terms of possibly degenerate elliptic partial differential equations. In this section, we review the results in Caselles et al. (2006), which are an extension to manifolds of the results in Caselles et al. (1998).

Let $(\mathcal{M}, g)$ be a compact, connected smooth two-dimensional surface in $\mathbb{R}^3$. As usual, given a point $\xi \in \mathcal{M}$, we denote by $T_\xi \mathcal{M}$ the tangent plane to $\mathcal{M}$ at the point ξ. Let $\mathcal{C}$ denote the family of continuous curves $\Gamma : [a, b] \to \mathcal{M}$ which are one-to-one in (a, b) and $\Gamma(a) = \Gamma(b)$. Let $\mathcal{D}$ denote the family of open subsets Ω of $\mathcal{M}$ such that the boundary of Ω, denoted by $\partial\Omega$ consists of a finite union of curves in $\mathcal{C}$. For each $\Omega \in \mathcal{D}$, let $C(\partial\Omega)$ be the set of continuous functions defined on $\partial\Omega$.

We shall consider an interpolation operator as a transformation E which associates with each $\Omega \in \mathcal{D}$ and each $\varphi \in C(\partial\Omega)$ a unique function $u = E(\varphi, \partial\Omega)$ defined on Ω satisfying the following set of assumptions:

- (A1) **Boundary Values**:

$$E(\varphi, \partial\Omega)|_{\partial\Omega} = \varphi \quad \text{for any } \Omega \in \mathcal{D} \text{ and } \varphi \in C(\partial\Omega).$$

  In other words, $E(\varphi, \partial\Omega)$ represents an interpolation or extension of φ into Ω.

- (A2) **Comparison Principle**:

$$E(\varphi, \partial\Omega) \leq E(\tilde{\varphi}, \partial\Omega)$$

  for any $\Omega \in \mathcal{D}$ and any $\varphi, \tilde{\varphi} \in C(\partial\Omega)$ with $\varphi \leq \tilde{\varphi}$.

- (A3) **Stability Principle**: Let $\Omega, \Omega' \in \mathcal{D}$, $\Omega' \subseteq \Omega$. Then

$$E(E(\varphi, \partial\Omega)|_{\partial\Omega'}, \partial\Omega') = E(\varphi, \partial\Omega)|_{\Omega'}$$

  holds for any $\varphi \in C(\partial\Omega)$. This principle means that no new application of the interpolation can improve a given interpolant.

- (A4) **Regularity Principle**: Let us briefly state the regularity axiom Caselles et al. (2006). In any case, the concepts and notation used here will be considered with more detail below in this section. Let ξ be a point on $\mathcal{M}$, $U \subseteq \mathbb{R}^2$ an open set containing 0, and $\psi : U \to \mathcal{M}$ be any coordinate system such that $\psi(0) = \xi$. Let $g_{ij}(x)$ and $\Gamma_{ij}^k(x)$ (indices i, j, k run from 1 to 2) denote, respectively, the coefficients of the

first fundamental form of $\mathcal{M}$ and the Christoffel symbol computed in the coordinate system ψ. For simplicity we shall denote by G the (symmetric) matrix $(g_{ij}(0))$ and by Γ^k the matrix formed by the coefficients $(\Gamma^k_{ij}(0))$, $i, j, k = 1, 2$. Let us denote by B_r the geodesic ball of radius r around ξ. Let $SM(2)$ be the set of symmetric 2×2 matrices. Let $A = (A^i_j)$ be a matrix such that $GA \in SM(2)$, and $p = (p^i) \in \mathbb{R}^2$. We shall use Einstein's convention that repeated indices are summed, and we denote by $(a, b) = a_i b^i$.

We can now state the regularity principle. For any quadratic polynomial $Q : U \to \mathbb{R}$ given by

$$Q(x) = \frac{1}{2} g_{ij}(0) A^i_l x^l x^j + g_{ij}(0) p^i x^j + c$$

$$(6.1)$$

$$= \frac{1}{2}(GAx, x) + (Gp, x) + c,$$

the operator E should satisfy

$$\lim_{r \to 0} \frac{E(Q \circ \psi^{-1}|_{\partial B_r}, \partial B_r)(\xi) - Q \circ \psi^{-1}(\xi)}{r^2/2} = F(A, p, c, \xi, G, \Gamma^k)$$

$$(6.2)$$

where F is a continuous function of its first argument. This requirement embodies several properties. First, it expresses that the interpolant of a quadratic polynomial near ξ may be locally expressed in terms of its elements A, p, c, the point ξ, and the metric tensor and Christoffel symbols. Since any smooth function u on U is given locally as a quadratic polynomial, this (together with the comparison principle) implies that the operator depends only on the first and second derivatives of u. Moreover, when combined with the comparison principle it permits to prove that the interpolation operator is intrinsic and the regularity axiom also gives the transformation properties of F when we change coordinates (see Theorem 6.4).

Let us notice that, as in Caselles et al. (2006), in (6.2) we could have written that the limit converges to a function $F(B, q, c, \xi, G, \Gamma^k)$ where $B = GA$, $q = Gp$. But this would mean only an equivalent change of notation. Following Caselles et al. (2006), we prefer to use this notation since it is more convenient in the proof of Theorem 6.2 (see Caselles et al. (2006)). Some further clarifying remarks will be given below.

- (A5) **Grey Scale Shift Invariance**:

$$E(\varphi + c, \partial\Omega) = E(\varphi, \partial\Omega) + c$$

for any $\Omega \in \mathcal{D}$, $\varphi \in C(\partial\Omega)$, $c \in \mathbb{R}$.

- (A6) **Linear Grey Scale Invariance**:

$$E(\lambda\varphi, \partial\Omega) = \lambda E(\varphi, \partial\Omega)$$

for any $\Omega \in \mathcal{D}$, $\varphi \in C(\partial\Omega)$, and any $\lambda \in \mathbb{R}$.

Let us describe the interpolation operators satisfying the above set of assumptions. To this goal, let us introduce some more notation which will also clarify the notation used in the regularity principle.
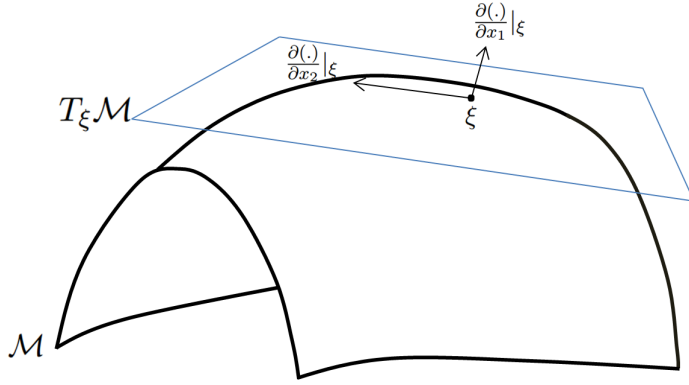
Figure 6.1: A point ξ on a manifold $\mathcal{M}$ and its tangent space at ξ.

For any $\xi \in \mathcal{M}$, we denote by $T_\xi\mathcal{M}$ the tangent space to $\mathcal{M}$ at the point ξ. By $T_\xi^*\mathcal{M}$ we denote its dual space. The scalar product of two vectors $v, w \in T_\xi\mathcal{M}$ will be denoted by $\langle v, w \rangle$, and the action of a covector $p^* \in T_\xi^*\mathcal{M}$, on a vector $v \in T_\xi\mathcal{M}$, will be denoted by (p^*, v). If $\psi : U \to \mathcal{M}$ is a coordinate system such that $\psi(0) = \xi$, and $g_{ij}(x)$ are the coefficients of the first fundamental form of $\mathcal{M}$ in ψ, we shall often write $g_{ij}(\xi')$ instead of $g_{ij}(x)$ where $x \in U$ is such that $\psi(x) = \xi'$. Then, if $v, w \in T_\xi\mathcal{M}$, we have $\langle v, w \rangle_\xi = g_{ij}(\xi)v^i w^j$, where v^i, w^i are the coordinates of v, w in the basis $\frac{\partial}{\partial x^i}|_\xi$ of $T_\xi\mathcal{M}$. Using this basis for $T_\xi\mathcal{M}$ and the dual basis on $T_\xi^*\mathcal{M}$, if $p^* \in T_\xi^*\mathcal{M}$, and $v \in T_\xi\mathcal{M}$, we have $(p^*, v) = p_i v^i$. Notice that we may write

$(p^*, v) = g_{ij}(\xi)p^i v^j$ where p^i are the coordinates of the vector p associated to the covector p^*. The relation between both coordinates is given by

$$p_i = g_{ij}(\xi)p^j, \quad \text{or} \quad p^i = g^{ij}(\xi)p_j, \tag{6.3}$$

where $g^{ij}(\xi)$ denotes the coefficients of the inverse matrix of $g_{ij}(\xi)$. By a slight abuse of notation, we shall write (6.3) as

$$p^* = Gp \quad \text{or} \quad p = G^{-1}p^*.$$

In this way $G : T_\xi \mathcal{M} \to T_\xi^* \mathcal{M}$. In the case that ψ is a geodesic coordinate system, the matrix G is the identity matrix $I = (\delta_{ij})$, and I maps vectors to covectors, i.e., $I : T_\xi \mathcal{M} \to T_\xi^* \mathcal{M}$. We shall denote by I^{-1} the inverse of I, mapping covectors to vectors.

Let us now clarify the notation used in (6.1). If $U \subseteq \mathbb{R}^2$, and $\psi : U \to \mathcal{M}$ is a coordinate system with $\psi(0) = \xi$, then $\psi \circ d\psi(0)^{-1} : U' \subseteq T_\xi \mathcal{M} \to \mathcal{M}$ is a new coordinate system. If we identify $T_0 U$ with $\mathbb{R}^2$ and $\{e_i\}$ denotes its canonical basis, then $e_i' = d\psi(0)e_i$ satisfy $\langle e_i', e_j' \rangle = g_{ij}(\xi)$. From now on, we shall use this identification, thus we shall interpret that any coordinate system around a point $\xi \in \mathcal{M}$ is defined on a neighborhood of 0 in the tangent space $T_\xi \mathcal{M}$.

We shall also use this coordinate system to express a bilinear map $\hat{A} : T_\xi \mathcal{M} \times T_\xi \mathcal{M} \to \mathbb{R}$. Indeed, if (A_{ij}) is the matrix of $\hat{A}$ in this basis, and $v, w \in T_\xi \mathcal{M}$, we may write $\hat{A}(v, w) = A_{ij}v^j w^i$. If $A^i_j = g^{ik}(\xi)A_{kj}$, then A^i_j determines a map, called $A : T_\xi \mathcal{M} \to T_\xi \mathcal{M}$ such that $\hat{A}(v, w) = \langle Av, w \rangle = (GAv, w)$. Observe that $GA : T_\xi \mathcal{M} \to T_\xi^* \mathcal{M}$. Observe also that our notation A^i_j already indicates that $A = (A^i_j)$ maps vectors to vectors, and this is the interpretation of the matrix argument A in (6.2). We shall identify matrices with maps.

As usual, we say that a linear map $C : T_\xi \mathcal{M} \to T_\xi^* \mathcal{M}$ is symmetric if $(Cv, w) = (Cw, v)$ for any $v \in T_\xi \mathcal{M}$, $w \in T_\xi \mathcal{M}$. From now on, we shall use the notation

$$SM(2) := \{A : T_\xi \mathcal{M} \to T_\xi^* \mathcal{M}, \text{ A is symmetric}\}.$$

Given the previous notations we can rewrite the Regularity priciple. Let $\Gamma_{ij}^k(x)$ the Christoffel symbol of $\mathcal{M}$ in the ψ coordinate system. Let us

denote B_r the geodesic ball of radius r around ξ. The operator E should satisfy:

$$\lim_{r \to 0} \frac{E(Q \circ \psi^{-1}|_{\partial B_r}, \partial B_r)(\xi) - Q \circ \psi^{-1}(\xi)}{r^2/2} = F(A, p, c, \xi, G, \Gamma^k).$$

The interpolation of a quadratic polynomial around ξ may be expressed in terms of its elements A, p, c, ξ, the metric tensor G and its Christoffel symbol. Any smooth function u is given locally as a quadratic polynomial, implies that the operator E depends only on its first and second derivative of u.

We consider a function u on $\mathcal{M}$ with gradient and Hessian $D_{\mathcal{M}}u$ and $D^2_{\mathcal{M}}u$, respectively. Considering $D_{\mathcal{M}}u$ as the covector of $\frac{\partial u}{\partial x^i}$ and $D^2_{\mathcal{M}}u$ the matrix $\frac{\partial^2 u}{\partial x^i \partial x^j} - \Gamma^k_{ij} \frac{\partial u}{\partial x^k}$. Using this notation $D^2_{\mathcal{M}}u(\xi) : T_\xi \mathcal{M} \times T_\xi \mathcal{M} \to \mathbb{R}$ is a bilinear map. We write $\nabla_{\mathcal{M}}u$ the vector of coordinates $g^{ij} \frac{\partial u}{\partial x^j}$. So that $|\nabla_{\mathcal{M}}u(\xi)|^2_\xi = \langle \nabla_{\mathcal{M}}u(\xi), \nabla_{\mathcal{M}}u(\xi) \rangle|_\xi$. Instead of writing $D_{\mathcal{M}}u$ and $\nabla_{\mathcal{M}}u$ to simplify the notation we are going to write Du and ∇u.

Given a function u on $\mathcal{M}$, let us denote by $D_{\mathcal{M}}u$ and $D^2_{\mathcal{M}}u$ the gradient and Hessian of u, respectively. In a coordinate system $D_{\mathcal{M}}u$ is the covector $\frac{\partial u}{\partial x^i}$, and $D^2_{\mathcal{M}}u$ is the matrix $\frac{\partial^2 u}{\partial x^i \partial x^j} - \Gamma^k_{ij} \frac{\partial u}{\partial x^k}$ which acts on tangent vectors. Thus, with this notation $D^2_{\mathcal{M}}u(\xi) : T_\xi \mathcal{M} \times T_\xi \mathcal{M} \to \mathbb{R}$ is a bilinear map, $\xi \in \mathcal{M}$. Let us write $\nabla_{\mathcal{M}}u$ the vector of coordinates $g^{ij} \frac{\partial u}{\partial x^j}$. Then $|\nabla_{\mathcal{M}}u(\xi)|^2_\xi = \langle \nabla_{\mathcal{M}}u(\xi), \nabla_{\mathcal{M}}u(\xi) \rangle_\xi$. To simplify our notation we shall write Du and ∇u instead of $D_{\mathcal{M}}u$, and $\nabla_{\mathcal{M}}u$. The vector field ∇u satisfies $\langle \nabla u, v \rangle_\xi = du(v)$, $v \in T_\xi \mathcal{M}$, du being the differential of u.

Now, we can recall here the results in Caselles et al. (2006) which are fundamental for our work. The main theorem of Caselles et al. (2006) shows that the interpolation functions $u = E(\varphi, \partial\Omega)$ are solutions of a partial differential equation. To simplify the presentation, we introduce the following notation

$$\Lambda_1(u, \xi) = D^2u \left(\frac{\nabla u}{|\nabla u|_\xi}, \frac{\nabla u}{|\nabla u|_\xi} \right)(\xi),$$

$$\Lambda_2(u, \xi) = D^2u \left(\frac{\nabla u}{|\nabla u|_\xi}, \frac{\nabla u^\perp}{|\nabla u|_\xi} \right)(\xi),$$

$$\Lambda_3(u, \xi) = D^2u \left(\frac{\nabla u^\perp}{|\nabla u|_\xi}, \frac{\nabla u^\perp}{|\nabla u|_\xi} \right)(\xi),$$

Definition 6.1. *Let* $(\mathcal{H}) : SM(2) \times \mathbb{R} \times (\mathcal{M})M \to \mathbb{R}$. *We shall say that* $\mathcal{H}(A, s, \xi)$ *is elliptic if* $\mathcal{H}$ *is a nondecreasing function of* A. *If* $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ *and we define* $\mathcal{H}(a, b, c, s, \xi) = \mathcal{H}(A, s, \xi)$, *we shall also say that* $\mathcal{H}(a, b, c, s, \xi)$ *is elliptic.*

Theorem 6.2. *Assume that the interpolation operator* E *satisfies* (A1), (A2), (A3), (A4), *and* (A5). *Then there exists a continuous function* $\mathcal{H} : SM(2) \times \mathbb{R} \times \mathcal{M} \to \mathbb{R}$, $\mathcal{H}(A, q, \xi)$, *which is nondecreasing in* A *such that if* $\Omega \in \mathcal{D}$, $\theta \in C(\partial\Omega)$, *and* $u = E(\theta, \partial\Omega)$, *then* u *is a viscosity solution of*

$$\mathcal{H}\left(\Lambda_1(u, \xi), \Lambda_2(u, \xi), \Lambda_3(u, \xi), |\nabla u|_\xi, \xi\right) = 0 \quad in \ \Omega, \tag{6.4}$$

satisfying the boundary data $u|_{\partial\Omega} = \theta$, *that is, for any* $\varphi \in C^\infty(\Omega)$ *with bounded derivatives such that* $u - \varphi$ *has a local maximum (minimum) at* ξ_0, *and* $\nabla\varphi(\xi_0) \neq 0$, *we get*

$$\mathcal{H}\left(\Lambda_1(\varphi, \xi_0), \Lambda_2(\varphi, \xi_0), \Lambda_3(\varphi, \xi_0), |\nabla\varphi|_{\xi_0}, \xi_0\right) \geq 0 \tag{6.5}$$

(respectively, ≤ 0*).*

Lemma 6.3. *Assume that the interpolation operator* E *satisfies* (A1)$-$(A5). *If, in addition, it satisfies* (A6), *then*

$$\mathcal{H}(\lambda a, \lambda b, \lambda c, |\lambda p|_\xi, \xi) = \lambda \mathcal{H}(a, b, c, |p|_\xi, \xi), \tag{6.6}$$

for any $a, b, c \in \mathbb{R}$, $\xi \in \mathcal{M}$, $p \in T_\xi\mathcal{M}$, $\lambda > 0$.

Let us finish this Section with some complementary information about $\mathcal{H}$. For simplicity, in the next proposition we shall not denote the argument ξ of $\mathcal{H}$.

Proposition 6.4. *i) If* $\mathcal{H}$ *does not depend upon its first or its third argument, then it also does not depend on its second argument. In other terms*

$$If \ \mathcal{H}(\alpha, \beta, \gamma, s) = \hat{\mathcal{H}}(\beta, \gamma, s), \ then \ \mathcal{H} = \hat{\mathcal{H}}(\gamma, s)$$
$$If \ \mathcal{H}(\alpha, \beta, \gamma, s) = \hat{\mathcal{H}}(\alpha, \beta, s), \ then \ \mathcal{H} = \hat{\mathcal{H}}(\alpha, s)$$

where $\alpha, \beta, \gamma, s \in \mathbb{R}$.

ii) If $\mathcal{H}$ *is differentiable at* $(0, 0, 0, 0)$ *then* $\mathcal{H}$ *may be written as* $\mathcal{H}(A, s) = tr(BA) + ds$ *where* B *is a nonnegative matrix with constant coefficients and* d *is a real constant.*

The case of the AMLE and Biased AMLE. A Comparison.

In this Chapter we focus on particular cases of equation (6.4) of Theorem 6.2 in previous chapter, namely, the AMLE and biased AMLE operators and discuss on their behavior. We illustrate it with a toy example. This proof of concept shows their capability to extend data to big empty regions. Additionally, this toy example allows us to show the different qualitative behavior of interpolation solution of the AMLE and of the biased AMLE.

7.1 Absolutely Minimizing Lipschitz Extensions (AMLE) and Biased AMLE

From previous chapter (see also Caselles et al. (2006)), if we assume that $\mathcal{H}$ is differentiable at $(0, 0, 0, 0)$ then we may rewrite Equation (6.4) as

$$a\Lambda_1(u, \xi) + 2b\Lambda_2(u, \xi) + c\Lambda_3(u, \xi) + d|\nabla u|_\xi = 0, \qquad (7.1)$$

where $a, c \geq 0$ and $ac - b^2 \geq 0$. If we take $a = c = 1$, $b = d = 0$ we recover the Laplace-Beltrami operator. If we choose $a = 1$, $b = c = d = 0$ we obtain the extension to two-dimensional manifolds of the so-called infinite Laplacian or AMLE, which stands for Absolutely Minimizing Lipschitz Extension. The authors of Caselles et al. (2006) also showed experimental results showing the potential applications of the AMLE model to interpolate data given on a set of curves or points on a surface.

The biased Absolutely Minimizing Lipschitz Extensions appeared in the axiomatic analysis of the interpolation operators in manifolds in Caselles et al.

(2006) (e.g., $a = 1$, $b = c = 0$ and $d = \beta$) although was not experimentally studied there (and was also unnoticed in Caselles et al. (1998)). It was also introduced in the context of tug-of-war games in Peres et al. (2010) where it was mathematically studied. It has been also been studied in Armstrong et al. (2009).

Let us recall that the AMLE was introduced by G. Aronsson in Aronsson (1967, 1968) and uniqueness of viscosity solutions was proved in Jensen (1993). For a survey we refer to Aronsson et al. (2004); Crandall (2008). On the other hand, the AMLE with variable coefficients was the object of study in Juutinen (1998) and more recently in Gómez and Rossi (2013).

As it is well known Caselles et al. (1998); Crandall et al. (2001); Aronsson et al. (2004), the AMLE is a cone interpolator and is able to interpolate the value at points. The biased AMLE offers an interesting alternative since the value at the points will expand and produce a smoother profile. This is of interest in applications to depth interpolation.

The authors of Peres et al. (2010) proved that the solution of the bi-ased AMLE in $\mathbb{R}^n$ with the Euclidean metric $d(x,y)$ is given by $u(x) = sign(\beta)(1 - e^{|\beta| d(x_0, x)})$. They refer to this solution as β-exponential cones. Figure 7.3 show the exponential cones obtained for the toy experiments (see also Figure 1 in Peres et al. (2010)). In general, d is a path distance between point x and the point at the center in x_0, and a continuous function is a solution of the biased AMLE if and only if it satisfies a certain comparison with β-exponential cones.

7.2 Biased AMLE

We consider the simple case where the Riemannian manifold $(\mathcal{M}, g)$ is given by $\mathcal{M} = \mathbb{R}^2$, or a domain Ω in $\mathbb{R}^2$, endowed with a metric g. Let $\theta \in C(\partial\Omega)$. Let us consider the problem

$$\Delta_{\infty, g} u + \beta |\nabla u|_\xi = 0 \quad \text{in } \Omega \tag{7.2}$$

coupled with the boundary condition

$$u|_{\partial\Omega} = \theta. \tag{7.3}$$

In (7.2) we have denoted

$$\Delta_{\infty, g} u := D_{\mathcal{M}}^2 u \left(\frac{\nabla u}{|\nabla u|_\xi}, \frac{\nabla u}{|\nabla u|_\xi} \right).$$

The equation (7.2) is called in the literature the biased AMLE. We consider equation (7.2) in the viscosity sense. Again, for convenience of notation, let us write

$$D^2_{\mathcal{M},ij}u(x) = \frac{\partial^2 u}{\partial x^i \partial x^j}(x) - \Gamma^k_{ij}(x)\frac{\partial u}{\partial x^k}(x).$$

Then, we can write (7.2) as

$$\frac{1}{|\nabla u(x)|^2_x}D^2_{\mathcal{M},ij}u(x)g^{i\alpha}(x)\frac{\partial u}{\partial x^\alpha}(x)g^{j\gamma}(x)\frac{\partial u}{\partial x^\gamma}(x) + \beta|\nabla u(x)|_x = 0 \quad (7.4)$$

where

$$|\nabla u(x)|^2_x = g^{ij}(x)\frac{\partial u}{\partial x^i}(x)\frac{\partial u}{\partial x^j}(x)$$

Let us describe a particular case which is of interest for us here. Assume that $g_{ij}(x) = h(x)\delta_{ij}$. In practice we assume that $h(x) = (\kappa_1 + \kappa_2|\nabla I(x)|^2)^p$ for some $p > 0$, where I is a given reference image. Then we may write (7.4) as

$$\frac{1}{|\nabla u(x)|^2}D^2_{\mathcal{M},ij}u(x)\frac{\partial u}{\partial x^i}(x)\frac{\partial u}{\partial x^j}(x) + \beta|\nabla u(x)|_x = 0 \quad (7.5)$$

Note that the factors $h(x)$ have been simplified in the numerator and the denominator in the first term of (7.5). They remain in the Christoffel symbols and in the modulus of the gradient in the second term.

Remark 7.1. *We can write also (7.4) as*

$$D^2_{\mathcal{M},ij}u(x)g^{i\alpha}(x)\frac{\partial u}{\partial x^\alpha}(x)g^{j\gamma}(x)\frac{\partial u}{\partial x^\gamma}(x) + \beta|\nabla u(x)|^3_x = 0 \quad (7.6)$$

Remark 7.2. *The AMLE is usually obtained as the limit of $p \to \infty$ of minimizers of the energy*

$$\int_\Omega |\nabla u|^p_g|g|^{1/2}\,dx \quad (7.7)$$

where $|g| = \det(g(x))$ satisfying specified Dirichlet boundary conditions. The case $p = 2$ corresponds to the Dirichlet integral. One could say that the work in Belyaev and Seidel (2005); Galic et al. (2015) is inspired by the Dirichlet integral and the case of a metric given by the structure tensor.

Let us illustrate this behavior using a toy example, created synthetically to be used as a proof of concept in the context of completion of sparse depth maps. The toy example consists of a constant depth map which contains
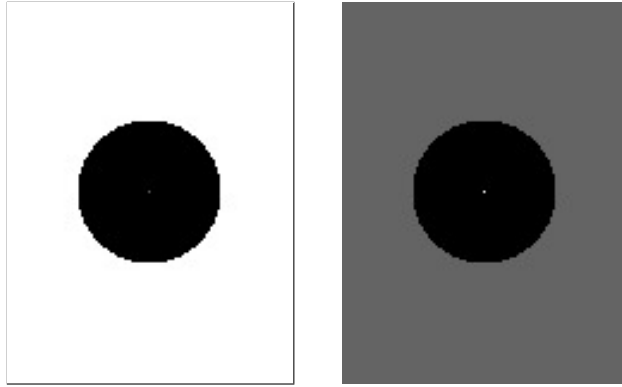
Figure 7.1: Left: Image of depth data. Outside the empty disc (the hole) the value is set to 255 and the only value at the center of the hole is set to 175. Right: Image of depth data. Outside the hole is set to 100 and the only value at the center of the hole is set to 175.

a circular hole (a disc of radius 30 pixels) representing a region lacking of depth data as we show in Figure 7.1 (a).

Actually, there is only one point in the disc with valid depth data, located at the center of the disc. We have performed this experiment twice: first, the constant depth value outside the disc is set to 255 and the only value at the center of the hole (the disc) is set to 175 (Figure 7.1 (a)); second, the constant depth value is set to 100 and the value at the center of the hole is set to 175 (Figure 7.1 (b)).

The result of both experiments is shown in Figure 7.2 and Figure 7.3. The AMLE operator produces an interpolation solution which is a cone interpolating from the boundary of the hole (the disc) to the valid point depth data located at the center of the disc. The biased AMLE operator produces an interpolation solution which is a higher order interpolating surface from the boundary of the hole (the disc) to the valid point depth data located at the center of the disc. Results were obtained using the euclidena metric.

By comparing qualitatively the solution of the AMLE and the solution of Biased AMLE for this toy example, we can interpret that the AMLE is a more suitable interpolator when extending depth date in images that can be approximated by planes between given values. On the contrary, for interpolating depth data in images with more complex profiles, the Biased AMLE results a more suitable interpolator.
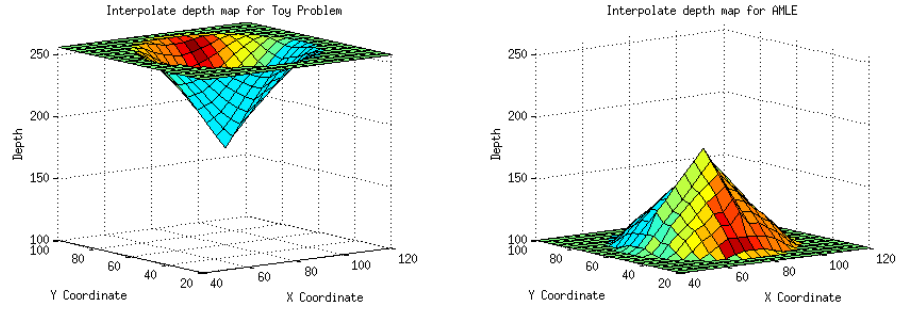
Figure 7.2: Left: AMLE interpolation solution when the constant value of the depth data outside the empty disc (the hole) is set to 255 and the only value at the center of the hole is set to 175. Right: AMLE interpolation solution when the constant value of the depth data outside the hole is set to 100 and the only value at the center of the hole is set to 175.
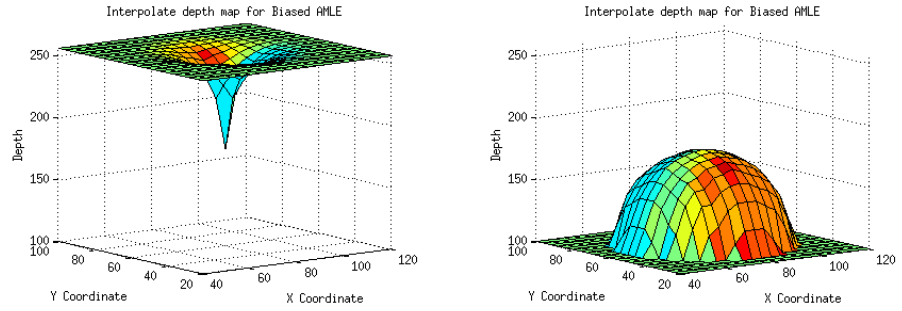
Figure 7.3: Left: Biased AMLE interpolation solution when the constant value of the depth data outside the empty disc (the hole) is set to 255 and the only value at the center of the hole is set to 175. Right: Biased AMLE interpolation solution when the constant value of the depth data outside the hole is set to 100 and and the only value at the center of the hole is set to 175. For both cases we consider $\beta = 1.2$

.

Numerical implementation

In this chapter we propose a numerical scheme for the biased AMLE partial differential equation (7.2). It builds on the convergent discretization of the infinit Laplacian PDE proposed by Oberman (2005a) (also studied in Manfredi et al. (2012), reinterpreting the discretization as a PDE on a graph). The algorithm turns out in an simple and fast explicit iterative scheme.

8.1 A numerical scheme for the biased AMLE

Let $I : \Omega \to \mathbb{R}$ be a given image. This image is our reference image, used to construct the metric g on $\Omega \subset \mathbb{R}^2$. In Chapter 7 we commented that we will consider a metric $g_{ij}(x) = h(x)\delta_{ij}$ with $h(x) = (\kappa_1 \|x\|^2 + \kappa_2 |\nabla I(x)|^2)^p$ for some $p > 0$, where I is a given reference image.

Let us now define a geodesic distance on the discrete image domain $\Omega \subset \mathbb{Z}^2$. We consider the discrete grid as a graph. For each two points x, y in the grid, let d_{xy} be its distance, defined as follows. If the two points are neighbors (that is, y is in the 4-neighborhood of x and x is in the $4-$neighborhood of y. Or, depending on the underlying discrete connectivity notion, both x and y are in its respective $8-$neighborhoods), its distance is given by

$$d_{xy} = \kappa_c |\vec{I}(x) - \vec{I}(y)|^2 + \kappa_x \|(x_1, x_2) - (y_1, y_2)\|^2, \tag{8.1}$$

where κ_c, κ_x are real constants. We have denoted the references image by $\vec{I}$ in order to stress that it is a color reference image. In other words, d_{xy} is a weighted sum of the square differences channel by channel of the

color values $\vec{I}(x)$ and $\vec{I}(y)$ with the euclidean distance between the points $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Later on, we will additionally, test another proposal for the distance that includes the depth map.

Given a curve $\{\gamma(i)\}_{i=0}^{m}$ in the discrete grid (where $\gamma(i)$ and $\gamma(i+1)$ are neighbors), we define the length of γ by

$$L_g(\gamma) = \sum_{i=0}^{m} d_{\gamma(i),\gamma(i+1)}. \tag{8.2}$$

Given any two points in the grid, then its geodesic distance d_{xy} is defined by:

$$d_{xy} = \inf\{L_g(\gamma) : \gamma \ is \ a \ curve \ \ joining \ x \ to \ y \ \}. \tag{8.3}$$

The distance d_{xy} is symmetric in x, y. Let us notice that this distance can be computed using Dijkstra algorithm. In the remainder of this section we give an algorithm for computing an approximation to d_{xy} and we provide an analysis of its accuracy with respect the exact d_{xy} (computed using Dijkstra algorithm).

8.1.1 Approximated discrete distance and a comparison with the exact one computed using Dijkstra algorithm
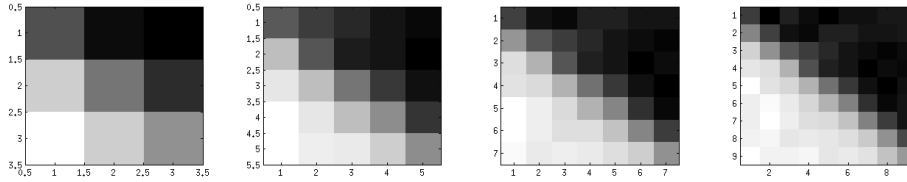
Figure 8.1: Squared neighborhoods of the reference image I centered at at pixel x of different radius.

Let us assume that we are given a gray level reference image I. Given any two pixels x and y, in our implementation, we compute its distance approximated d_{xy} by (8.1). This approximation let us to reduce the processing time. We present here an experimental comparison with respect the exact d_{xy} given by (8.3) and computed using Dijkstra algorithm.

To this aim, let us fix a pixel x in the image domain and compute the exact and the approximated distance d_{xy} for all y belonging to the four squared neighborhoods, of four different radius, shown in Figure 8.1. In each neighborhood in Figure 8.1 w The computed geodesic distances from the central pixel x to every pixel in the neighborhood is displayed in Figure 8.2, plotted a surface. The exact geodesic distances (8.3) are shown in the left column while the approximated distances (8.1) are shown is the right column. Let us notice that the plotted surfaces appear to be similar and we claim that the error committed in the solution of the biased AMPLE equation produced by this approximation of the geodesic distance will be small.

8.2 Numerical implementation of Biased AMLE

Given a point x in the grid, let $\mathcal{N}(x)$ denote any neighborhood of x made of a connected set of points on the grid, *except* the point x itself.

Following Manfredi et al. (2015) the positive eikonal operator on a graph is defined by

$$\|\nabla u(x)\|_x^+ = \max_{y \in \mathcal{N}(x)} \frac{u(y) - u(x)}{d_{xy}}. \tag{8.4}$$

The negative eikonal operator on a graph is defined by

$$\|\nabla u(x)\|_x^- = \min_{y \in \mathcal{N}(x)} \frac{u(y) - u(x)}{d_{xy}}. \tag{8.5}$$

The discrete infinity Laplacian corresponds to (Oberman (2005b))

$$\Delta_{\infty,g} u(x) = \frac{\|\nabla u(x)\|_x^+ + \|\nabla u(x)\|_x^-}{2}. \tag{8.6}$$

We propose the following discrete version of the biased AMLE equation (7.2):

$$\frac{\|\nabla u(x)\|_x^+ + \|\nabla u(x)\|_x^-}{2} + \beta \left| \|\nabla u(x)\|_x^+ \right| = 0 \tag{8.7}$$

with $\beta > 0$. Thus, in order to numerically solve (8.7), we discuss on the sign of $\|\nabla u(x)\|_x^+$. If the eikonal operator is positive, (8.7) writes

$$\frac{\|\nabla u(x)\|_x^+ + \|\nabla u(x)\|_x^-}{2} + \beta \|\nabla u(x)\|_x^+ = 0, \tag{8.8}$$
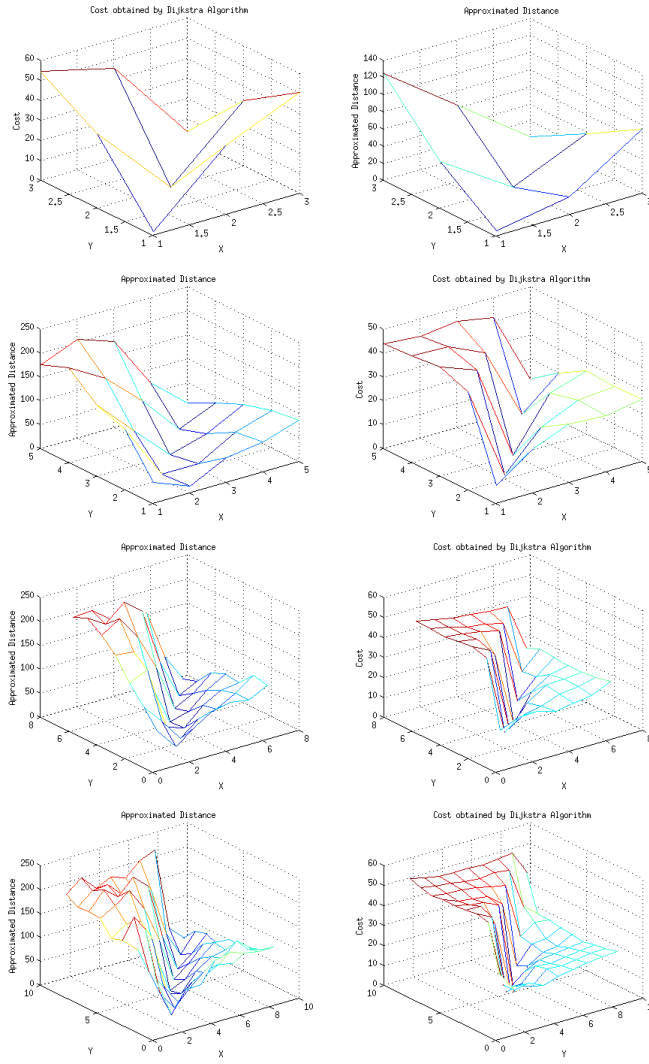
Figure 8.2: Exact and approximated geodesic distances computed from the central pixel of the neighborhood to every pixel in the neighborhood. Each row corresponds to different radius of the neighborhood, as shown in Figure 8.1. Left column: exact geodesic distances (8.3) Right column: approximated geodessic distances.

and, if the eikonal operator is negative, (8.7) writes

$$\frac{\|\nabla u(x)\|_x^+ + \|\nabla u(x)\|_x^-}{2} - \beta\|\nabla u(x)\|_x^+ = 0. \tag{8.9}$$

Both expresions (8.8) and (8.9) can be written as

$$\frac{\|\nabla u(x)\|_x^+ + \|\nabla u(x)\|_x^-}{2} + \beta\text{sign}(\|\nabla u(x)\|_x^+)\|\nabla u(x)\|_x^+ = 0. \qquad (8.10)$$

Alternatively, the discrete version of the biased AMLE (8.10) can be written as

$$\beta_+\|\nabla u(x)\|_x^+ + \beta_-\|\nabla u(x)\|_x^- = 0, \qquad (8.11)$$

where $\beta_+ > \beta_-$ (actually, $\beta_+ = \frac{1}{2} + \beta\text{sign}(\|\nabla u(x)\|_x^+)$ and $\beta_- = \frac{1}{2}$). Notice that, if $\beta_+ = \beta_-$, then (8.11) would be a multiple of the infinity Laplacian. In that case, the scheme is known to be convergent Oberman (2005b) (and the proof extends to open sets in a Riemann surface).

Finally, let us compute the explicit scheme. Given x, let $y, z \in \mathcal{N}(x)$ be such that

$$\|\nabla u(x)\|_x^+ = \frac{u(y) - u(x)}{d_{xy}}$$

and

$$\|\nabla u(x)\|_x^- = \frac{u(z) - u(x)}{d_{xz}}.$$

By introducing these expressions into (8.11) we obtain

$$u(x) = \frac{\beta_+ d_{xz} u(y) + \beta_- d_{xy} u(z)}{\beta_+ d_{xz} + \beta_- d_{xy}}. \qquad (8.12)$$

The numerical scheme for the discrete biased AMLE is

$$u^{k+1}(x) = \frac{\beta_+ d_{xz} u^k(y) + \beta_- d_{xy} u^k(z)}{\beta_+ d_{xz} + \beta_- d_{xy}}, \qquad k = 0, 1, ... \qquad (8.13)$$

We will show experiments considering the several posibilities for β_+ and β_- (e.g., $\beta_+ = 2$, $\beta_+ = 1$) and considering several neighborhoods $\mathcal{N}(x)$. Let us stress again that if $\mathcal{N}(x)$ coincides with the 4-neighborhood of x, or the 8$-$neighborhood of x, then the exact distances d_{xy} are simply computed by (8.1). If we take a larger neighborhood we use our approximation to (8.3).

This type of scheme was studied in Oberman (2005b) for the AMLE. Let us also point out the connections with the approach in Almansa et al. (2002) although the authors did not consider the case of manifolds.
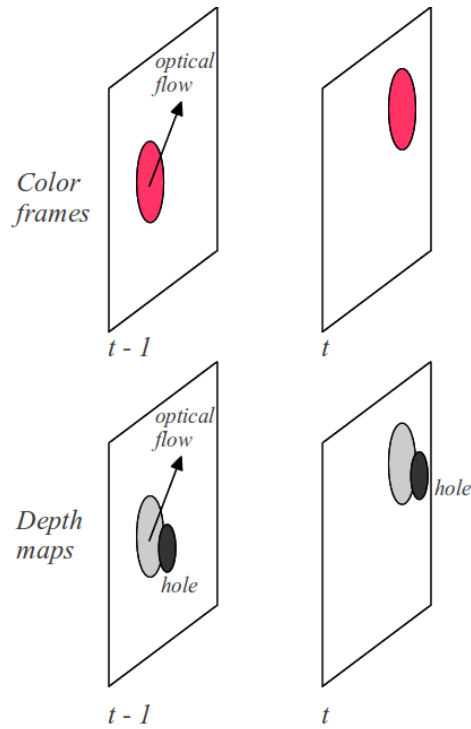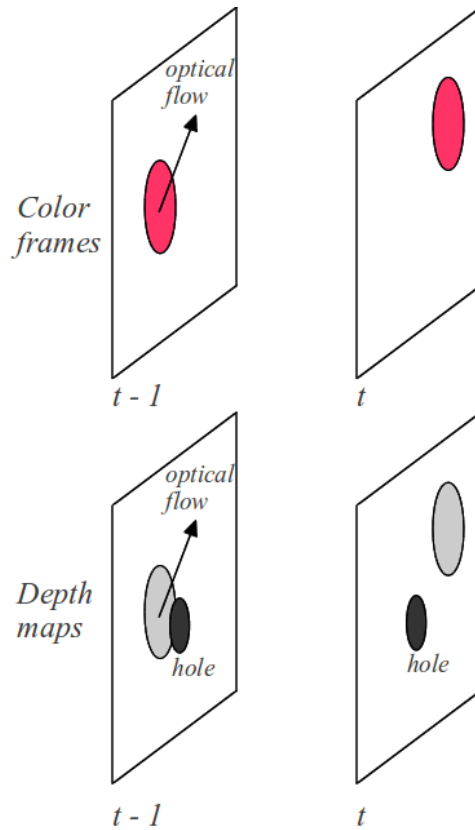
Figure 8.3: Color frames and depth map for a video sequence. In the color images a red balloon moves from left to right. We show the optical flow as a black arrow in frame $t-1$. We show the depth map for the red balloon that moves and we show a hole in the depth map. The hole has the same motion of the object. No additional information can be obtained compensating the depth map by the optical flow.

8.3   Extension to video

We have extended the AMLE and the biased AMLE to also handle temporal information. Let us consider two consecutive frames u_{t-1} and u_t of a depth video sequence. Suppose that the depth image u_{t-1} has holes or missing data. We propose to consider the available depth data in u_{t-1} and the available depth data in the consecutive frame u_t in the video sequence as well as the reference video.

We assume that there is an available optical flow computed from the video sequence. We use this optical flow to compensate the depth map u_t. With this new information we construct a new interpolation mask considering the information in u_t and in u_{t-1}.

Figure 8.4: Color frames and depth map for a video sequence. In the color images a red balloon moves from left to right. The black arrow represents the optical flow. We show the depth map for the red balloon that moves and we show a hole in the depth map. In this example the hole has a different motion of the object, it means that additional information can be obtained compensating the depth map by the optical flow.

We show in Figure 8.3 a red balloon that moves from left to right. We also show the depth map of the balloon. There is a hole in the depth map that moves solidary with the balloon. By compensating u_t, we don't have new depth information.

In Figure 8.3 we show a slightly different situation. The red ballon moves from let to right but the hole in the depth map does not move solidary with the object. When we compensate the depth data in u_t we have new depth information that helps to complete the depth data in u_{t-1}.

We have at our disposal a depth map u_t of a depth video sequence and also a color reference image $\vec{I}_t$. The depth image can present lack of information. In the place where lack of information is present (holes) we create a binary mask (if we have depth data $mask = 0$ else $mask = 1$). This mask represents the interpolation domain for a completion algorithm.

The presented procedure described above is a tool to brig new information from depth map in u_t to time the instant $t - 1$. With this new information we modify the interpolation binary $mask$. With the new information we fill holes of the binary mask reducing the amount of points to be interpolated, improving the quality of the interpolated depth map.

We apply these ideas completing depth video sequences in the next chapter.

Experimental Results

In this chapter, we present some experimental results obtained with the discussed interpolation operators. We also provide a comparison between the biased AMLE, the AMLE and the bilateral filter in the context of depth upsampling and enhancement. On the other hand, we present experiments for the extension of these interpolation operators to the video case. We evaluated the proposed methods on different database namely Middlebury, ENPEDA EISATS and KITTY.

9.1 Sampling depth images

We evaluate on upsampling experiments the performance of AMLE, biased AMLE and bilateral filter. To this goal we have sampled depth data. We have used depth images of the Middlebury dataset. We also consider the color reference image of the Middlebury dataset. In Figure 9.1 we show the color reference image and the depth image.

We show in Figure 9.1 the images we used to evaluate the performance AMLE, biased AMLE and bilateral filter. In 9.1 Figure (a) we show the reference image for the image teddy. In (b) we show the depth image (ground truth) for teddy. In (c) we show the reference images for image baby. In (d) depth image (ground truth) for Baby. In (b) and (d) we superpose to the original image a red square, the idea is to enlarge this area for posterior processing.

We took the depth data in Figure 9.1 and we sampled it with differents space sample rate. We show in Figure 9.2 the enlarged region of the teddy

(a)   (b)

(c)   (d)
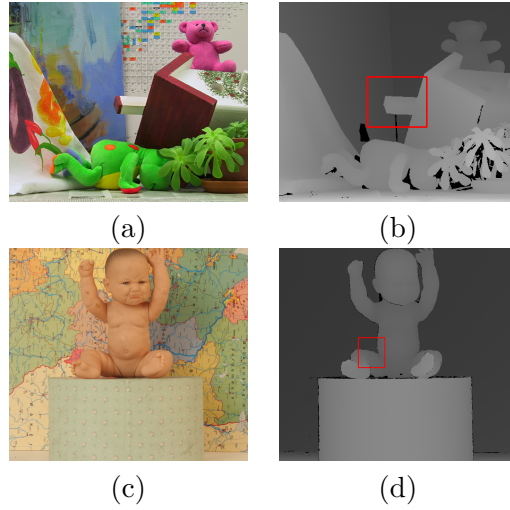
Figure 9.1: Reference color image and depth image for teddy and baby of Middlebury database.

sampled depth image using differents sample rate. We sample the depth data in a regular regular grid as can be seen in Figure 9.2 (a), (b), (c), (d).

(a)   (b)

(c)   (d)
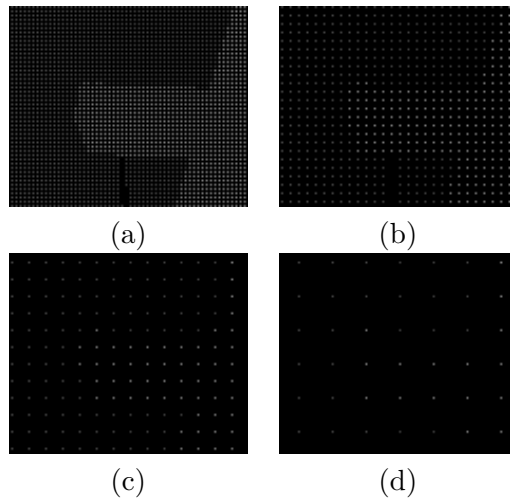
Figure 9.2: Sampled teddy depth image. In (a) there is 1 sample every other 4 pixels square. (b) 1 sample for every other 16 pixels square. (c) 1 for every other 64 pixels square. (d) 1 sample for every other 256 pixels square.

We present in Figure 9.3 sampled version of Baby depth image.
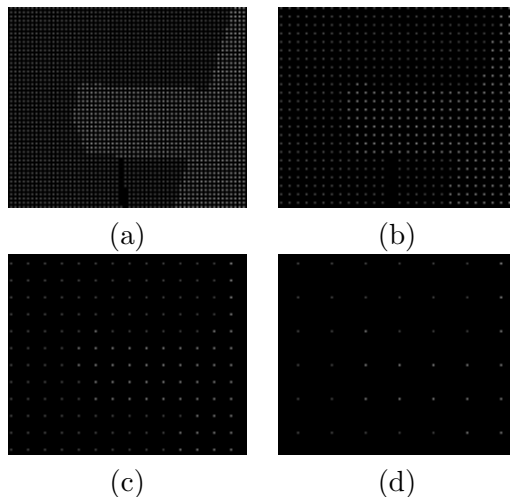
Figure 9.3: Sampled version of the baby depth image. In (a) there is 1 sample for every other 4 pixels square. (b) 1 sample for every other 16 pixels square. (c) 1 sample for every other 64 pixels square. (d) 1 sample for every other 256 pixels square.

We show in Figure 9.3 the enlarged region of the Baby sampled depth image using differents sample rate. We sampled the depth data in a regular regular grid as can be seen in Figure 9.3 (a), (b), (c), (d).

We have compared three differents upsampling methods: AMLE, biased AMLE and bilateral filter. These methods were tested in two sampled depth images (teddy and baby) of Middlebury database Baker et al. (2011).

Table 9.1: Parameters for AMLE filter to complete sampled baby depth image

baby			
radius	1	2	5
κ_x	5	10	100
κ_y	5	10	100
κ_c	50	100	
κ_d	50	100	
β^+	1		
β^-	1		

We show in Table 9.2 the parameter values for biased AMLE method used in sampled depth image.

Table 9.2: Parameters for biased AMLE filter to evaluate sampled depth image

baby			
radius	1	2	5
κ_x	5	10	100
κ_y	5	10	100
κ_c	50	10	100
κ_d	50	100	
β^+	1.1	1.2	1.5
β^-	1		

We show in Table 9.3 the parameters value for bilateral filter to complete sampled depth image.

Table 9.3: Parameters for sampled baby depth image

baby	min		max
radius	1	2	5
κ_x	5	10	100
κ_y	5	10	100
κ_c	50	100	-
κ_d	50	100	-

9.1.1 teddy evaluation

We have used the robust mean absolute error(mae) to evaluate the results of each interpolation method:

$$mae = \frac{1}{N} \sum_{i \in \Omega}^{N} |gt(x) - u(x)| , \qquad (9.1)$$

where gt is the ground truth, u is the interpolated depth image, N is the number of points to interpolate, Ω es the interpolation domain. This mae has been chosen because of this robustness against outliers.

Our first goal is to validate the good performance of biased AMLE and AMLE in the upsampling task. In Figure 9.4 we show results completing sampled teddy depth image. We show in Figure 9.4 (a) result obtained by AMLE. (b) and (c) result obtained by biased AMLE with different β

mae=1.5068

(a)

mae=1.4727

(b)

mae=1.4614

(c)

mae=1.5089

(d)
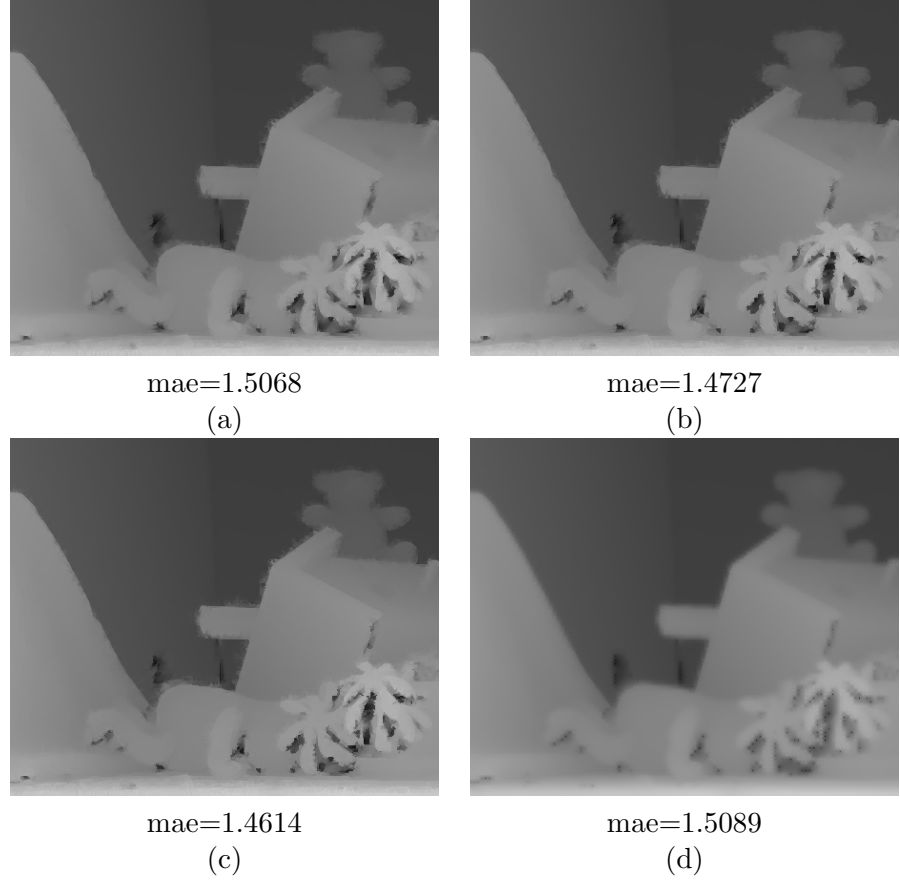
Figure 9.4: Results of Upsampling task for an image of Middlebury database. (a) Results obtained by AMLE. (b) Biased AMLE $\beta = 1.1$. (c) Biased AMLE $\beta = 1.2$. (d) Bilateral filter.

parameter. (e) bilateral filter. We observe the minimum mae error was obtained by biased AMLE with $\beta = 1.2$. The result in (c) looks more defined in the edges compared with AMLE in (a) and bilateral filter in (d).

We have performed upsampling experiments using the AMLE filter, biased-AMLE filter and bilateral filter. we have used the original depth map as a ground truth to evaluate their performance. We see in Table 9.4 that the biased AMLE outperform AMLE and bilateral filter in upsampling task. We took the average value of results for every sampled version of teddy depth image avoiding variations of the results. For this image results obtained by AMLE, biased AMLE and bilateral filter are very similar.

Table 9.4: Results obtained by AMLE filter, biased AMLE filter and by bilateral filter for subsample teddy depth image

Size	AMLE	biased AMLE $\beta^+ = 1.1$	biased AMLE $\beta^+ = 1.2$	bilateral
Filtered 16×16	2.5749	2.5798	2.5988	2.5535
Filtered 8×8	2.0304	2.0010	2.0019	2.0935
Filtered 4×4	1.5068	1.4727	1.4614	1.5089
Filtered 2×2	1.0884	1.0891	1.0748	1.0288
Average	1.8001	1.7857	**1.7842**	1.7962

9.1.2   baby evaluation

Table 9.5: Results obtained by AMLE filter, biased AMLE filter and by bilateral filter for subsampled baby depth image

Size	AMLE	biased AMLE $\beta^+ = 1.1$	biased AMLE $\beta^+ = 1.2$	bilateral
Filtered 16×16	1.2593	1.2589	1.2620	2.0485
Filtered 8×8	0.9618	0.9503	0.9516	1.6447
Filtered 4×4	0.7486	0.7312	0.7298	1.2091
Filtered 2×2	0.5614	0.5596	0.5548	0.9106
Average	0.8828	0.8750	**0.8746**	1.4532

Similar to the evaluation of teddy depth image we performed upsampling experiments using the AMLE filter, biased-AMLE filter and bilateral filter for Baby depth image. we have used the original depth map as a ground truth. We see in Table 9.5 that the biased AMLE outperform AMLE and bilateral filter in upsampling task. We also took the average value of the mae to avoid variations of the results. For every sampled version of Baby depth image AMLE and biased AMLE outperform the bilateral filter. The size of the objects in Baby image is greater than objects in teddy. In this image biased AMLE shows its capability to extend the depth information to huge regions.

9.1.3   Temporal extension experiment

We have extended to temporal domain the biased AMLE. We have considered the procedure explained in section 8.3. We evaluate the extended biased AMLE filter in a video sequence of ENPEDA EISATS database. In Figure 9.5 we present example frames of the selected video.
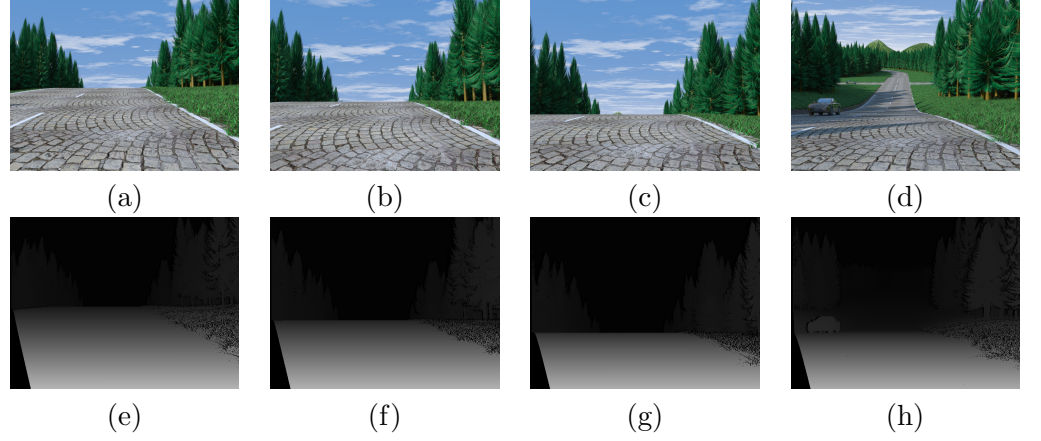
Figure 9.5: Images of the ENPEDA EISAT video sequence database. (a) frame 1 of the video sequence, (b) frame 25, (c) frame 50, (d) frame 100, (e) synthetic depth for frame 1, (f) frame 25 (g) frame 50 (h) frame 100.

In Figure 9.5 we present some frames of the synthetic video sequence. In 9.5 (a) we show frame 1, (b) frame 25, (c) frame 50 and (d) frame 100 in (e), (f), (g), (h) we show the corresponding depth map for frame 1, 25, 50,100 respectively.

We added artificial holes to the depth data of the synthetic video. We show in Figure 9.6 the original depth data and the depth data plus holes.

In Figure 9.6 we added holes to the original data. The holes are rectangular holes located in random positions. Our idea is to produce this kind of perturbation in the depth video sequence to complete the lack of information using biased AMLE operator. Taking into account that we have at our disposal the optical flow of the video sequence. We have two possibilities to create perturbations in the depth data: i) add random holes to every frame of the sequence. ii) propagate the holes using the optical and create a tube in the video. In Figure 9.7 we show random holes added to every frame of the depth video sequence and holes propagated using the optical flow.

(a)

(b)
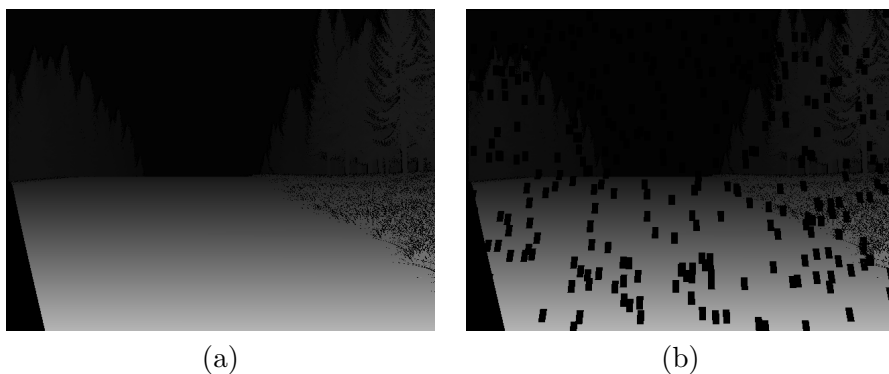
Figure 9.6: Original depth data and holes added to the depth data. (a) original image. (b) holes added to the original data.

(a)

(b)

(c)

(d)
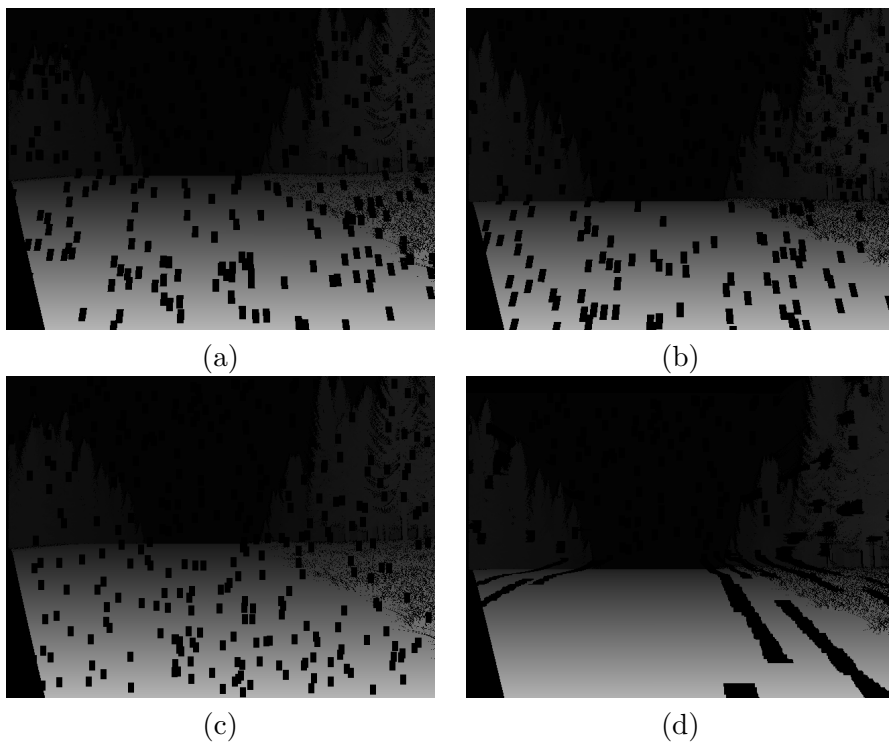
Figure 9.7: Depth data perturbed with random holes. (a) frame 1 perturbed with random holes. (b) frame 25 with random holes. (c) frame 1 with random holes. (d) holes propagated uisng the optical flow.

In Figure 9.7 in (a) and (b) we show random holes added to every frame of the depth video sequence. In (c) and (d) the holes are propagated using the optical flow. We see that the holes becomes lines due to in the initial frame (c) the hole is a part of a tube of the video.

We have two perturbed video sequences: random holes and propagated holes video sequences.

We evaluate the biased AMLE and time extended biased AMLE in this video sequences. We show results obtained by the biased AMLE Table 9.6. We show results obtained by the time extended biased AMLE Table 9.7.

Table 9.6: Results obtained by biased AMLE filter in EISATS video sequence

	biased AMLE $\beta^+ = 1.1$	biased AMLE $\beta^+ = 1.2$	biased AMLE $\beta^+ = 1.5$
Propagates holes	1.529	1.624	2.197
Random holes	1.071	1.295	1.998

Table 9.7: Results obtained by time extended biased AMLE filter in EISATS video sequence

	biased AMLE $\beta^+ = 1.1$	biased AMLE $\beta^+ = 1.2$	biased AMLE $\beta^+ = 1.5$
Propagated holes	1.441	1.492	2.043
Random holes	1.026	1.180	1.550

We observe in Table 9.6 and in Table 9.7 that for Propagates holes video sequences and for Random Holes video sequences the time extended biased AMLE outperforms the biased AMLE. We conclude that the addition of time information improves the depth completion. Additionally, we observe that for larger β values larger error is obtained.

9.2 Completion of depth data in KITTI database

9.2.1 KITTI database

In this section, we present some results on completion of sparse data for images for images of the KITTI database. The KITTI database consists
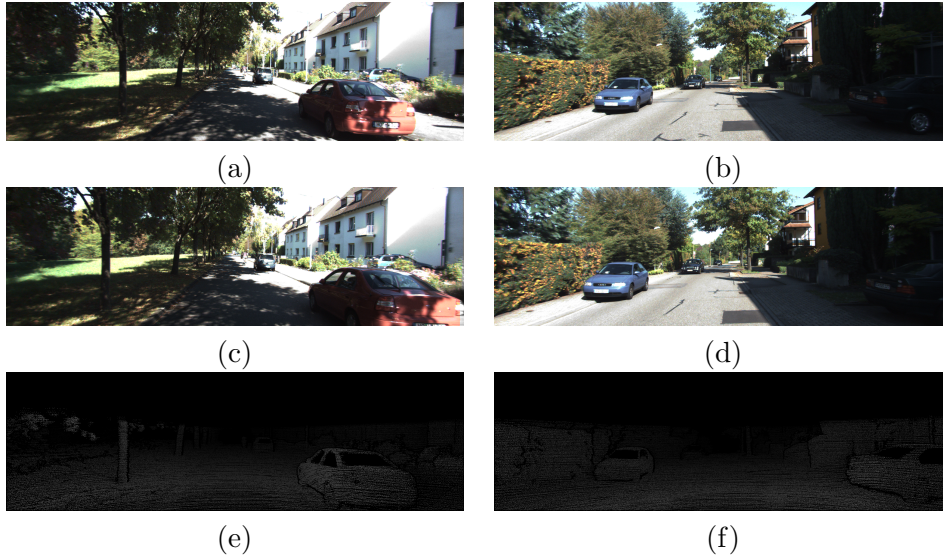
Figure 9.8: Images of the KITTI video sequence database. (a) first frame of sequence 11, (b) first frame of sequences 32, (c) second frame of sequence 11, (d) second frame of sequence 32, (e) depth image for first frame sequence 11, (f) depth image for first frame sequence 32

of 194 real scenes recorded form a moving vehicle. The data set consists of 194 pairs of consecutive color images. The database includes a ground truth provided by a laser scanner.

We show in Figure 9.8 examples of KITTI database.

Figure shows results obtained using the AMLE, biased AMLE and bilateral filter. Let us notice that the biased AMLE produces a result to the ground truth reference image(see Figure (b)). Finally let us notice that the bilateral filter result shown in Figure (c) is very blury.

The obtained results in Figure 9.9 shows that AMLE and biased AMLE can we can extend the depth map to huge region (a) and (b). The bilateral filter take weighted average and the final result is very blur (c). In the other hand biased AMLE extend the depth taking into account the geometry of the image (b).
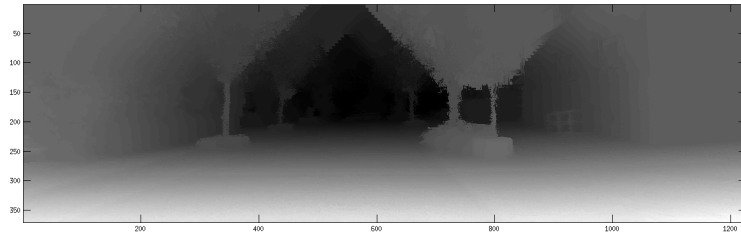
(a)
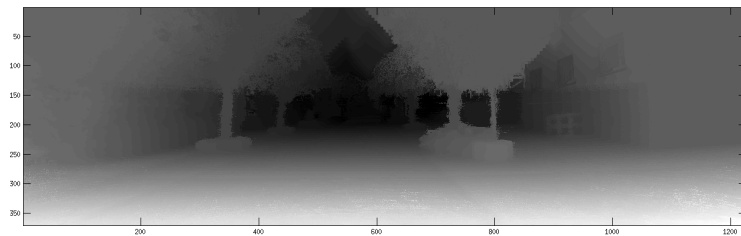
(b)
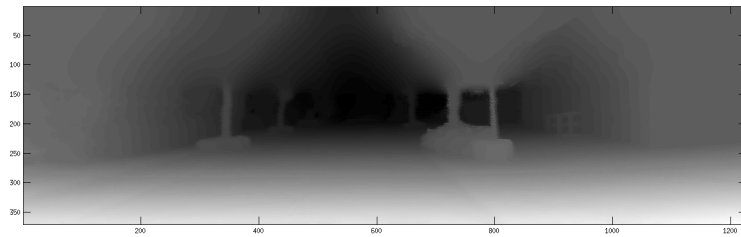
(c)

Figure 9.9: Results for an example from KITTI database. From top to bottom: results obtained by AMLE (a), biased AMLE (b), Bilateral filter (c).

Comparison between biased AMLE and Gradient based non-local bilateral filter

This chapter presents a comparison between the different depth enhanced methods proposed and studies in tis manuscript, in particular, the gradient based non-local filter presented in part I of this document and the biased AMLE studied in Part II.

We have evaluated both methods in the experiments we already performed in this Part I and in Part II.

10.1 Experiments

10.1.1 Two planes experiment

The experiment consists of completion of a two plane depth image. The depth data is only known outside of the hole in the center of the depth image. This hole defines the interpolation area as we show in Figure 10.1(a). We have at our disposal the reference image 10.1(b).

In Figure 10.2 we compare the profiles of the solutions obtained by gradient-based non-local bilateral filter and biased AMLE($\beta = 1.1$).

In the Figure 10.2 we show a profile of the extended depth image in the interpolation area. The Figure shows the results of the gradient-based non-local bilateral filter (blue crosses) and the results by biased AMLE(red and

(a) (b) (c) (d)
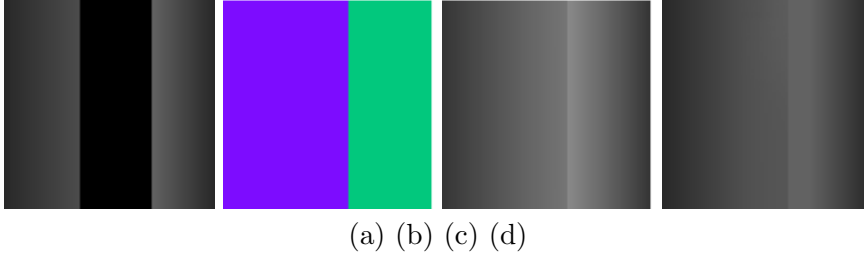
Figure 10.1: In this Figure we present: synthetic depth data (the black region indicates the interpolation area), reference image, result using gradient-based, result using biased AMLE with $\beta = 1.1$.
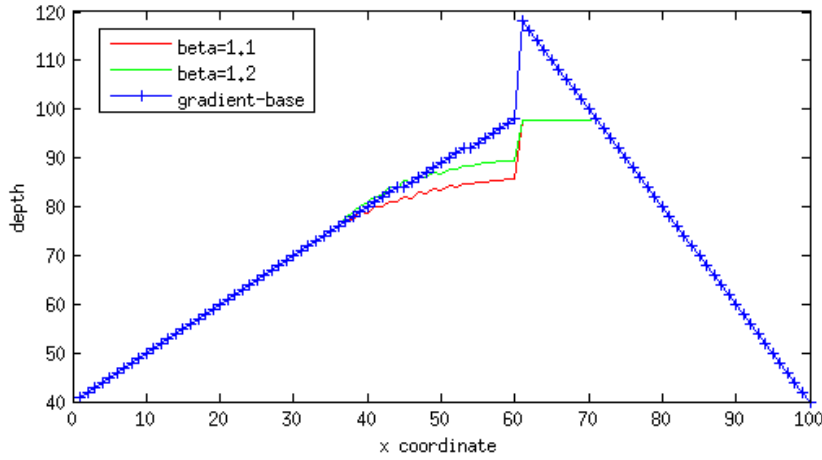
Figure 10.2: Profile of the result for the two planes synthetic image. We show 3 curves: output of biased AMLE with $\beta = 1.1$, biased AMLE with $\beta = 1.2$ and output of gradient-based non-local bilateral filter.

.

green lines). We see that the result for $\beta = 1.1$ presents staircase effect and in the left side, the method is not able to estimate the slope of the plane. With $\beta = 1.2$ (green line) the result improve the estimation of the slope in the left side, but in the right side is not well estimated.

10.1.2   Examples with an image from Middlebury

We present results for upsampling experiment. We have already presented in previous sections results of the biased AMLE for teddy image. In Figure

(a) (b)

(c) (d)
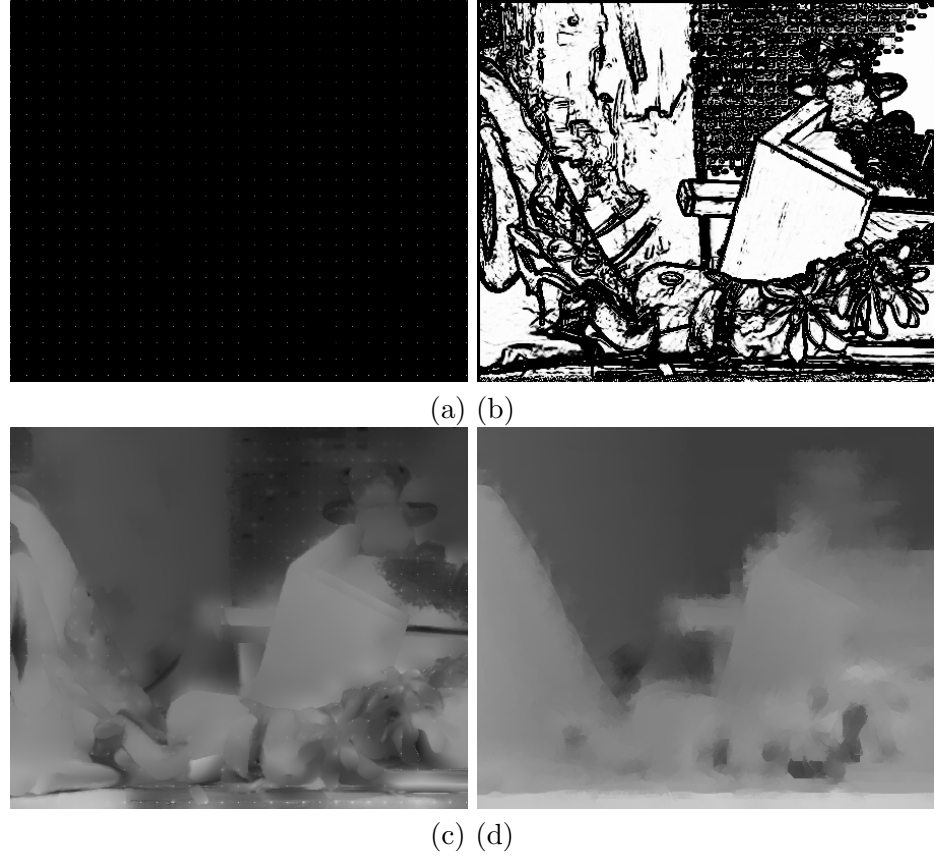
Figure 10.3: Results for upsampling teddy. (a) subsampled image teddy depth image. (b) edge map for teddy. (c) Result obtained by gradient-based method. (d) Result obtained by biased AMLE.

10.3 we show results for gradient-based non-local bilateral filter and biased AMLE. We have upsampled teddy image considering filtered data with a downsample factor 16 (in every other two hundred fifty-six square-pixel there is one pixel).

In the Figure 10.3 we show the data to upsample, the gradient mask needed for gradient-based algorithm, the output the gradient-based algorithm and the result of biased AMLE. We have evaluated the mae for both method. As we previously reported for biased AMLE we have a mae= 2.5798 and for gradient based a mae= 4.1986.

10.1.3   Examples with real image

We present the result of both algorithms for an image that presents large areas with lack of depth information. The depth is not reliable in many areas of the depth image.

We present in Figure 10.4 the reference image, the initial depth data and the edge map. Note the large empty areas to the right and to the left border of the depth image.

The Figure 10.5 shows results obtained by gradient-based(a) and by biased AMLE (b). We see that output of the gradient-based in large empty areas is blurry and the method can not extend the information till the end of the region. Biased AMLE is precise in large empty areas and extend the disparity as a flat area.
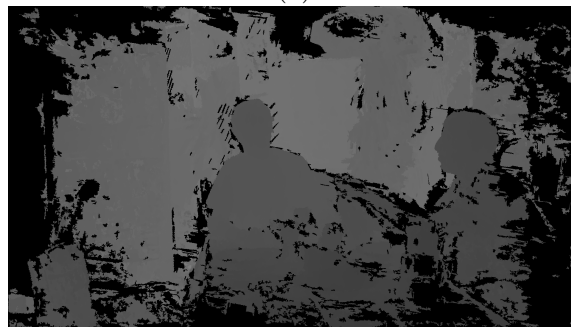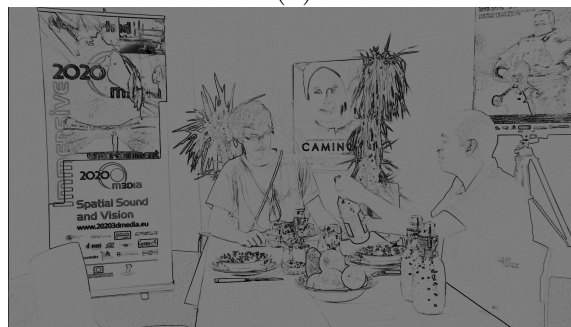
Figure 10.4: (a) Color reference image. (b) Initial depth data obtained by a depth sensor. (c) Edge map for gradient-based non-local bilateral filter algorithm.

(a)

(b).

Figure 10.5: Results obtained by gradiente-based(a) and biased AMLE(b).

Conclusions

In the case we have a depth image that presents large region with lack of information, the axiomatic stated by Caselles et al. (2006) lets us construct an interpolator to complete these large empty regions.

The interpolator is computed a ratio between two terms: a sum of function values weighted by distances and a sum of distances. This simple operation made the operator easy to implement, simple and fast.

The numerical implementation consider many very strong approximation that experimentally we have shown that their are valid.

We have evaluated the Biased AMLE in differents experiments: umsampling, temporal extension. Upsampling experiments. Results show that the AMLE and Biased AMLE outperform the bilateral filter. In upsampling experiments, in the case of filtered depth images, the biased AMLE outperforms the AMLE. That show that the inclusion of the gradient in the biased AMLE helps to estimate better the depth data surface.

We have extended the biased AMLE to temporal domain. We added a new term to the distance function. This new term considers the available depth information in the next image. The use of the optical flow helps to improve the performance of the depth interpolation in video sequences.

We have compared the biased AMLE and the gradient-based non-local bilateral filter in different experiments that show their potential and differences.

As Future work we propose constructing a joint interpolator between the gradient-base non-local bilateral filter and biased AMLE. We propose to use a balance term between these two methods. This balance term should

use the gradient-base algorithm in the case of depth images that can be approximated by locally planes and in the case that the image presents large areas of missing data, complete them using the biased AMLE.

Illumination Changes and Large Displacement Robust Optical Flow

The optical flow problem

12.1 Introduction

The apparent movement of pixels in a sequence of images is usually called the optical flow. Optical flow computation is one of the most challenging problems in computer vision, specially in real scenarios where occlusions and illumination changes occur. It has many applications, including autonomous flight of vehicles, insertion of objects on video, slow camera motion generation, video compression and many more.

In video compression, the optical flow estimation helps to remove temporal data redundancy and therefore to attain high compression ratios. In video processing, optical flow estimation is used, e.g., for deblurring, noise suppression or motion-compensated 3D sampling structure conversion Stiller and Konrad (1999). Ruffier and Franceschini (2005), Kendoul et al. (2005) present several autopilot systems controling aerial autonomous vehicles which are based on optical flow. In the work of Ruffier and Franceschini (2005), a bioinspired autopilot was developed to control an autonomous aircraft. Their system is able to perform complex maneuvers such as automatically taking off or landing. In Kendoul et al. (2005) a joint optical flow, depth and self position estimation system was developed to autopilot an autonomous aircraft. They use a low resolution camera and an inertial sensor. The complete system is able to work in real-time.

In order to estimate the flow field that represents the motion of points in two consecutive frames of a video, most of the optical flow methods are grounded on the optical flow constraint. This constraint is based on the *brightness constancy assumption* which states that the brightness or inten-

sity of objects remains constant from frame to frame along the movement of the objects.

Let us consider two consecutive image frames, I_0 and I_1, of a video sequence, where $I_0, I_1 : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^2$ is the image domain (which is usually assumed to be a rectangle in $\mathbb{R}^2$). The aim is to estimate a 2D motion field $u : \Omega \to \mathbb{R}^2$, the optical flow, such that the image points $I_0(x)$ and $I_1(x + u(x))$ are observations of the same physical scene point. In other words, the brightness constancy assumption writes:

$$I_0(x) = I_1(x + u(x)) \tag{12.1}$$

for x in Ω. Let us assume that the displacement $u(x) = (u_1(x), u_2(x))$ is small enough to be valid the following linearized version of the brightness constancy assumption:

$$I_0(x) = I_1(x) + \nabla I_1(x) \cdot u(x), \tag{12.2}$$

where ∇I_1 denotes the gradient of I_1 with respect to the space coordinates. This equation can be rewritten as:

$$\partial_t I_1 + \nabla I_1(x) \cdot u(x) = 0, \tag{12.3}$$

where $\partial_t I_1$ denotes $I_1(x) - I_0(x)$. Equation (12.3) is usually called the *optical flow equation* or *optical flow constraint*.

The optical flow constraint equation (12.3) is only suitable when the partial derivatives can be correctly approximated. This is the case when the motion field is small enough or the images are very smooth. However, in the presence of large displacements, these conditions are not typically preserved, and it is common to replace it with a nonlinear formulation Meinhardt-Llopis et al. (2013).

$$I_0(x) - I_1(x + u(x)) = 0, \tag{12.4}$$

for $x \in \Omega$.

Neither (12.3) nor (12.4) can be solved pointwise since the number of parameters (the two components $u_1(x), u_2(x)$ of $u(x)$) to be estimated is larger than the number of equations. On the other hand, occlusions produce lack of correspondence between some points in the image sequence. Occluded pixels include pixels of an image frame which are covered by the movement

of object in the following frame. For those occluded pixels, there is no a reliable optical flow. To challenge these problems, a variational approach could be used and compute the optical flow u by minimizing the following energy or error measure,

$$E(u) = \int_{\Omega} |I_0(x) - I_1(x + u(x))|^2 \, dx. \tag{12.5}$$

However, (12.5) is an ill-posed problem which is usually challenged by adding a regularity prior. Then, the regularization term added to the energy model allows to define the structure of the motion field and ensures that the optical flow computation is well posed. Horn and Schunck (1981) proposed to add a quadratic regularization term. Actually, the work of Horn and Schunck (1981) was the first which introduced variational methods to compute dense optical flow. An optical flow u is estimated as the minimizer of the following energy functional

$$E(u) = \int_{\Omega} |I_0(x) - I_1(x + u(x))|^2 \, dx + \int_{\Omega} |\nabla u|^2 \, dx, \tag{12.6}$$

where $\nabla u = (\nabla u_1, \nabla u_2)$ denotes the gradient of the optical flow. This L^2 regularity prior does not cope well with motion discontinuities and other regularization terms have been proposed, Nagel and Enkelmann (1986); Black and Anandan (1996); Brox et al. (2004a); Zach et al. (2007); Werlberger et al. (2009); Sun et al. (2010a); Werlberger et al. (2010); Krähenbühl and Koltun (2012); Xu et al. (2012); Chen et al. (2013); Sánchez et al. (2014); Zimmer et al. (2011); Strekalovskiy et al. (2014); P.Palomares et al. (2015); Ranftl et al. (2014); Sun et al. (2014). Although the original work of Horn and Schunk reveals many limitations (e.g., the computed optical flow is very smooth and sensitive to the presence of noise), it has inspired many proposals. In order to cope with large displacements, optimization typically proceeds in a coarse-to-fine manner (also called a multi-scale strategy). The variational setting for estimating the optical flow will be briefly reviewed in Section 12.2.

Let us now focus on the data term in (12.6), $E_D(u) = \int_{\Omega} |I_0(x) - I_1(x + u)|^2 \, dx$. The brightness constancy constraint assumes that the illumination of the scene is constant over time (in other words, over the image sequence) and that the image brightness of a point remains constant along its motion tra-

jectory. Therefore, changes in brightness are only due to diferent objects and different movements.

However, this assumption is flawed in realistic scenarios, where illumination changes appear as well as occlusions due to the relative motion between objects in the scene or the camera movement. Also, shadows or light reflexions that appear and move in the image sequence can also make the brightness constancy assumption to fail. This is why in this work we consider an alternative to the classical brightness constancy constraint which will provide robustness against additive illumination changes.

Before finishing this introduction, let us briefly recall that a problem is called ill-posed if its solution either does not exist or it is not unique. We have observed that the optical flow equation (12.3) (or (12.4)) is an ill-posed problem: it cannot be solved pointwise as there is a unique equation and two unknowns u_1 and u_2. The optical flow equation (12.3) can be rewritten as:

$$\nabla I \cdot u = -\partial_t I, \tag{12.7}$$

where we have forgotten about the subindex and the point x. By considering a local orthonormal basis $\{e_1, e_2\}$ of $\mathbb{R}^2$ on the directions of ∇I and $\nabla I^\perp$, the optical flow u can be expressed in this basis, $u = u_p e_1 + u_\perp e_2$, where $u_p e_1$ is the projection of u in the gradient direction and $u_\perp e_2$ is the projection of the optical flow u in the direction perpendicular to the gradient. Then, equation (12.7) could be formally written as

$$\nabla I \cdot (u_p e_1 + u_\perp e_2) = \|\nabla I\| u_p = -\partial_t I. \tag{12.8}$$

This equation can be solved for u_p, $u_p = -\frac{\partial_t I}{\|\nabla I\|}$, if $\|\nabla I\| \neq 0$. In other words, the only component of the optical flow that can be determined from (12.3) is the component parallel to the gradient direction. This indeterminacy is called the *aperture problem.*

12.2 Optical flow estimation

Most accurate techniques that address the motion estimation problem are based on the formulation of the optical flow estimation in a variational setting. Energy-based methods are called global methods since they find correspondences by minimizing an energy defined on the whole image (as the

minimizing problem (12.6)). They provide a dense solution with subpixel accuracy, and are usually called dense optical flow methods.

In Horn and Schunck (1981), a seminal and highly influential work was presented. It estimates a dense optical flow field based on two assumptions: the brightness constancy assumption and a smooth spatial variation of the optical flow. They proposed the following functional:

$$E(u) = \int_\Omega |\partial_x I u_1 + \partial_y I u_2 + \partial_t I|^2 \, dx + \alpha^2 \int_\Omega |\nabla u_1|^2 + |\nabla u_2|^2 \, dx, \quad (12.9)$$

where α is a real parameter which controls the influence of the smooth term. This functional is convex and has a unique minimizer. However, the computed optical flow is very smooth and do not preserve discontinuities of the optical flow. This is also the case for (12.6), which can be considered a variant of (12.9).

After the initial work in Horn and Schunck (1981), a lot of approaches that focus on accuracy have been developed. These works focus on the use of robust estimators, either in the data or smoothness terms, to be able to deal with motion discontinuities generated by movement of different objects or by occlusions (e.g., Brox et al. (2004a); Black and Anandan (1996); Sun et al. (2010a); Zach et al. (2007)). For the data term, L^2 or L^1 dissimilarity measures have been used as well as more advanced data terms Brox et al. (2004a); Stein (2004); Müller et al. (2011). For the smoothness term, isotropic diffusion, image-adaptive, isotropic diffusion with non-quadratic regularizers, anisotropic diffusion (image or flow-adaptive) or the recent non-local regularizers have been proposed Nagel and Enkelmann (1986); Black and Anandan (1996); Brox et al. (2004a); Zach et al. (2007); Werlberger et al. (2009); Sun et al. (2010a); Werlberger et al. (2010); Krähenbühl and Koltun (2012); Xu et al. (2012); Chen et al. (2013); Sánchez et al. (2014); Zimmer et al. (2011); Strekalovskiy et al. (2014); P.Palomares et al. (2015); Ranftl et al. (2014); Sun et al. (2014). However, these methods may fail in occlusion areas due to forced, but unreliable, intensity matching. The problem can be further accentuated if the optical flow is smoothed across object boundaries adjacent to occlusion areas.

12.2.1 Robust motion estimation

Normally assumptions as brightness constancy and smooth space variation of optical flow are violated in real images. Advanced robust optical flow

Figure 12.1: From left to right: Two consecutive frames of the Backyard sequence from Middlebury dataset and the occlusion layer (*red*) superimposed on the first frame.

methods are developed with the goal to perform well even when violations of the optical flow assumptions are present Igual (2006).

The model (12.9) to estimate the optical flow penalizes high gradients of u and therefore doesn't allow discontinuities of u. This model is highly sensible to noise in the images and also to outliers. The functional can be modified in order to allow discontinuities of the flow field by changing the quadratic data term to a L^1 term and also changing the L^2 regularization term. In Zack et al. (2007) the authors present a novel approach to estimate the optical flow that preserves discontinuities and it is robust to noise. In order to compute the optical flow $\mathbf{u} = (u_1, u_2) : \Omega \to \mathbb{R}^2$ between I_0 and I_1, the authors propose to minimize the energy

$$E(u) = \int_\Omega (\lambda |I_0(x) - I_1(x + u(x))| + |\nabla u_1| + |\nabla u_2|) \, dx, \qquad (12.10)$$

including robust data attachment and regularization terms (namely, the Total Variation of $\mathbf{u}$) with a relative weight given by the parameter $\lambda > 0$. This variational model is usually called the *TV-L1 formulation*. The use of L^1 type-norm measures has proven a good performance in front of L^2 norms to preserve discontinuities in the flow field and offers an increased robustness against noise and illumination changes. We will extend in Chapter 13 (Section 13.1) the model (12.10) to jointly compute the optical flow and occlusions. We will model and estimate an occlusion mask identifying the occluded pixels, i.e. pixels that are visible in I_0 but not in I_1.

Occlusions are a challenging problem in the estimation of optical flow. Occluded regions in the current image of the video sequence are defined as the set of pixels which become hidden by occluding (moving) objects in the next image, as illustrated in Figure 12.1

These pixels have no corresponding points in the next image and motion is not observable at their locations. Therefor, methods may fail in occlusion areas due to forced, but unreliable, intensity matching. The problem can be further accentuated if the optical flow is smoothed across object boundaries adjacent to occlusion areas. Hence, the occlusion issue must be carefully addressed to ensure a reliable and accurate computed optical flow. Some methods implicitely deal with occlusion by using terms robust norms in the data term while others do an explicit occlusion handling.

A first step towards taking into account occlusions was done by jointly estimating forward and backwards optical flow in Alvarez et al. (2007). The authors argue that at non-occluded pixels forward and backward flows are symmetric. Thus, the occlusion is determined by introducing into the formulation an error measure that assesses, for each pixel, the consistency between the forward and backward flows. Intensity matching is still forced at occluded pixels. Optical flow estimation and occlusion detection are decoupled. In Ince and Konrad (2008), the authors propose a formulation that computes optical flow and implicitly detects occlusions, extrapolating optical flow in occluded areas. Occlusions are determined again by assessing the consistency of the forward and backward flows for each pixel. This cue is used to penalize the intensity matching accordingly. Thus, the method does not force matching at occluded pixels. The extrapolation mechanism in the occlusion areas is based on anisotropic diffusion and uses the underlying gradient to preserve optical flow discontinuities. Another joint approach for optical flow and occlusion detection was developed in Xiao et al. (2006). This work proposes a two step updating scheme. The first step updates the flow field based only on the data and occlusion cues, given by the mismatch in the intensity value between the two images. The second step performs a flow diffusion process using a bilateral filter that locally adjusts filter strength by means of the occlusion cue.

Layered approaches Wang and Adelson (1994) allow to realistically model occlusion boundaries. In order to do this one has to correctly compute the relative order of the surfaces. Performing inference over the combinatorial range of possible occlusion relationships is challenging. A recent work that explicitly models occlusions and depth ordering can be found in Sun et al. (2010b). The authors present a method in which a visibility function is estimated for each layer. Spatial and temporal constraints are imposed to these functions in order to ensure layer consistency. These functions are used to penalize the intensity matching functions correspondingly. The results obtained are very good but the computational load to minimize the

associated energy is high.

Other authors try to obtain occlusion boundary information by means of an operator directly applied to the computed motion. In Thompson et al. (1985), the authors argue that occlusion boundaries can be detected by assessing the points at which the flow changes rapidly. Discontinuities in the optical flow correspond in fact to zeros in the Laplacian fields of each motion component along the direction perpendicular to the occlusion boundary. In Sand and Teller (2008a) video motion is represented as a set of particles. As the authors point out, the divergence of the motion field can be used to distinguish between different types of motion areas. Schematically, the divergence of a flow field is negative for occluded areas, positive for dis-occluded, and near zero for the matched areas. Taking this into account, the authors define an intensity matching term that is weighted by a function depending on the divergence of the motion. At each iteration of the optimization procedure, the motion field is filtered, similarly to Xiao et al. (2006), with a bilateral filter that depends on the divergence of the motion. The latter idea is used in Sun et al. (2010a) in order to perform a robust median filtering of the motion. In another context, Corpetti et al. (2002) analyzes the problem of estimating the motion of fluids. They use a divergence-curl regularizer to be able to deal with large concentrations of vorticity and divergence.

In Xu. et al. (2010) a method to estimate occlusions is used. They consider the fact that multiple points mapped by the optical flow to the same point in the target image (collision) are likely occluded. The authors count the pixels of the reference image which are mapped to the same point in the target image using forward warping. To avoid noise in the occlusion estimation the authors use a small Gaussian filter.

A method to estimate occlusion is presented in A. Ayvaci et al. (2012). They suppose Lambertianian surfaces and also that the illumination is static. They state the occlusion estimation as a variational problem and simplify the functional to a sequence of convex problems which are solvable by using an efficient numerical scheme. The proposed method works for any number of independently moving objects. They present results in Middlebury database.

On the other hand, in real video sequences the brightness constancy assumption is frequently violated in practice due to changes in the illumination sources of the scene, shadows, noise in the acquisition process, specular reflections or large and complex deformation. A common counter-example

consists in moving the light source of an immobile scene, producing brightness variations without motion of any objects. Robustness against it would be desirable. Small variations in the brightness can be dealt with a criterion that is invariant under brightness value changes. The gradient of the image is robust to additive illumination changes in the images Brox et al. (2004b), and the gradient constancy assumption writes:

$$\nabla I_0(x) - \nabla I_1(x + u(x)) = 0 \tag{12.11}$$

This equation can be included as a new data term in a variational energy in order to compute the optical flow u Xu. et al. (2010).

The optimization strategy in variational optical flow is usually based on local optimization in conjunction with a coarse-to-fine method. Indeed, variational methods are embedded in a pyramidal coarse-to-fine approach which consists in an iterative refinement of the motion field in each scale. This strategy consists in creating a multiscale pyramid, estimating the optical flow at coarser scales and then refining the solution at finer scales Mémin and Pérez (1998); Brox et al. (2004b). However, while traditional methodology work well in cases where the small structures move more or less the same way as larger scale structures, the approach fails if the displacement of the object in the scene is greater than the size of the object itself. In recent years this topic has been tackled in interesting approaches. In Brox et al. (2009), a method for large displacements is proposed that performs region-based descriptor matching. This method estimates correctly large displacement but it can match outliers. Steinbruecker and Pock (2009) also propose a method in order to tackle large displacement. The methodology performs well in real images with large displacements but it presents a lack of subpixel accuracy.

The gradient constancy assumption and the robust L^1 norm are incorporated in an optical flow method in Xu. et al. (2010). To tackle large displacement they incorporate matchings of SIFT features computed between the images of the sequence. The fusion between the matchings of SIFT features and optical flow estimation is performed using graph cuts.

Weinzaepfel et al. (2013) presents a variational approach which incorporates sparse matches computed with what the authors call Deep Matching algorithm, which they adapt to optical flow estimation. The matching algorithm is based on a six layer-stage interleaving convolution and max-pooling, similar to deep learning networks. The integration of deepmatchings to the

optical flow estimation is performed adding a matching term to the variational model and solving it by a coarse-to-fine methodology.

SparseFlow finds sparse pixel correspondences by means of a matching algorithm and these correspondences are used to guide a variational approach to obtain a refined optical flow. SparseFlow matching algorithm uses an efficient sparse decomposition of pixels' surrounding patch as a linear sum of those found around candidate corresponding pixels Timofte and Van Gool (2015). The matching pixel dominating the decomposition is chosen. The pixel pair matchings in both directions (forward-backward) are used to refine the optical flow estimation.

In R. Kennedy and C. J. Taylor (2015) a successful method to compute optical flow is proposed which includes occlusion handling and additional temporal information. The images are divided into discrete triangles and this allows them to naturally estimate the occlusions which are then incorporated into the optimization algorithm to estimate an optical flow. They define the concept "Inertial Estimate" of the flow. Combining this concept and classifiers to fusion optical flow they reach some improvements in the final results.

D. Fortun et al. (2015) proposes a method to compute optical flow which aims to tackle large displacements, motion detail and occlusion estimation. The method consists of two stages. In the first stage they supply dense local motion candidates. They estimate affine motion models over a set of size-varying patches combined with patch-based pairing. They experimentally demonstrate that the motion vector estimated that way provides at least one accurate motion vector for each pixel. The second step consists on a discrete optimization algorithm which selects candidates at each pixel while ensuring piecewise smoothing of the resulting flow. Their method is evaluated in MPI Sintel Butler et al. (2012) database achieves state-of-the-art results.

In Chapter 13 (see Section 13.2.3) we propose a pointwise confidence measure able to decide whether an optical flow estimation is reliable. Our optical flow confidence measure allows to achieve better overall results by improving the optical flow obtained by the variational approach we propose in Section 13.1 and 13.2 with an optical flow obtained with sparse methods in those unreliable regions. Confidence measures are used in the recent literature to validate the optical flow field by estimating its correctness at each point. We finish this chapter by reviewing some of the existing proposals. Traditional confidence measures used in the literature are build from:

- Local analysis of the input images. The simplest one is the magnitude of the gradient of the first image $\|\nabla I_0\|$ justified by the assumptions that we expect a higher accuracy in textured than in flat regions. The motion of highly textured regions in general is well estimated and the magnitude of the gradient of I_0 can be interpreted as a measure of texture of a region.

- Distance to the nearest edges. Image edges may occur with motion boundaries, the higher the distance from edges, the lower the chase of occlusion O. M. Aodha et al. (2013).

- A measure of error in optical flow estimation based on the values $|I_0(x) - I_1(x + u)|$.

- Combinations of eigenvalues of the structure tensor.

- Bootstrap resampling: a computational statistical inference technique based on repeating the optical flow calculation several times for different randomly chosen subsets of pixel contributions Kybic and Nieuwenhuis (2011)

- Intermediate results or parameters of a particular OF estimation method. For instance, Bruhn and Weickert (2006) uses the local contribution to the total energy being minimized to identify locations where model assumptions are not valid and assignes low confidence to them.

The work of Patras et al. (2007) presents an statistical confidence measure of the motion vectors estimated by a block matching. The matching of blocks is expressed through a probability framework as a maximum likelihood estimation scheme. The confidence measure is expressed as a function of a posteriori probability and the likelihood of the estimated optical flow.

In C. Kondermann et al. (2008) a confidence measure for optical flow is presented. The confidence measure relies on test theory and statistics of the flow field. The confidence measure is applicable to any optical flow algorithm. The authors perform comparison with other existing confidence measures.

In O. M. Aodha et al. (2013), a supervised learning based method is presented to estimate a confidence value for each optical flow field at each pixel and decide if an optical flow algorithm is likely to fail in their estimation. The feature vector is constructed using temporal and spatial magnitudes.

Using this confidence value they can fusion four differents optical flow algorithms.

Proposed model

Optical flow problem concerns with the estimation of the apparent motion between two consecutive images of a video sequence. As reviewed in previous Chapter 12, most adopted strategies to solve the optical flow problem use a variational approach. In those vatiational models, the optical flow computation is stated as an energy minimization problem where the energy has, in general, two terms; namely, the data term and the regularization term.

In this chapter, we propose a variational model for joint optical flow and occlusion estimation, which is adapted for both gray and color image sequences and is able to handle illumination changes as well as large displacements. To better explain the different contributions, we have chosen to present it progressively. First, we propose a joint minimization problem to estimate both optical flow and occlusions while preserving discontinuities of the flow. Our data term is based on the brightness constancy constraint for the case of gray sequences and on the color constancy constraint for color sequences. The color constancy assumption or constraint states that the color of the pixels do not change along the displacement of the object Brox et al. (2004b). The proposed energy model incorporates information that allows to detect occlusions. This information is based on the divergence of the flow and the energy favors the location of occlusions on regions where this divergence is negative. Assuming that occluded pixels are visible in the previous frame, the optical flow on non-occluded pixels is forward estimated whereas is backwards estimated on the occluded ones.

As mentioned in previous chapter, the optical flow constraint presents some drawbacks: often gray or color constancy assumption is violated due to il-

lumination changes, shadows or reflexions, as well as due to occlusions that appear when objects or the camera move. On the other hand, most of the variational models for optical flow computation use the coarse-to-fine strategy to be able to handle large displacement of the objects in images Stoll et al. (2012). This strategy consists in creating a multiscale pyramid, estimating the optical flow at coarser scales and then refining the solution at finer scales Mémin and Pérez (1998); Brox et al. (2004b); Meinhardt-Llopis et al. (2013). However, the coarse-to-fine strategy is unable to handle large displacements of small objects that move differently from their surroundings. In this chapter we propose and analise a methodology to handle all these drawbacks.

The remainder of this chapter is organized as follows. In Section 13.1, we present a model that jointly computes the optical flow and occlusions. To simplify, the model is presented for gray level sequences. Then, in section 13.2 we extend it to a method which able to handle illumination changes as well as large displacements. Both sections include the proposed optimization method. Finally, Section 13.3 is devoted to present the numerical algorithm.

13.1   A Model to Jointly Compute Optical Flow and Occlusions

In order to construct a joint optical flow and occlusion estimation model, we build up from ideas behind existing approaches and the fact that the divergence of the field can be used to detect occlusion areas. We model the occluded pixels by using a binary occlusion mask identifying the pixels that are visible in a frame, but not in the following frame. Our proposal uses three consecutive frames: the first frame, the central frame and the last frame. We assume that pixels of the central frame that are not visible in the last frame, are visible in the first frame.

The method is based on the minimization of an energy made of three terms. The first term considers the L^1 norm in the brightness constancy assumption, the second term is based on Total Variation of the flow and the third term characterizing the occlusions considers the divergence of the flow multiplied by the estimated occlusion mask. The energy is strongly nonlinear and the terms are decoupled by introducing auxiliary variables. Then the energy is minimized by means of a numerical scheme which is based on: i) a dual formulation of the Total Variation, ii) an efficient point-wise thresholding step, and iii) a primal-dual algorithm to compute occlusion mask.

Let us introduce the proposal. Given two consecutive image frames $I_0, I_1 : \Omega \to \mathbb{R}$, with $\Omega \subset \mathbb{Z}^2$, the authors in Zach et al. (2007) proposed the following variational discontinuity preserving model in order to compute the optical flow $u = (u_1, u_2) : \Omega \to \mathbb{R}^2$ between I_0 and I_1:

$$E(u) = \int_\Omega \lambda \, |I_0(x) - I_1(x + u(x))| \, dx + \int_\Omega (|\nabla u_1| + |\nabla u_2|) \, dx, \quad (13.1)$$

with $\lambda > 0$. This model considers the L^1 norm in the brightness constancy assumption as data fidelity term and the total variation of u as the regularization term.

Based on the model in (13.1) we present a joint optical flow and occlusions estimation model. In Chapter 12 we have observed that the divergence of the motion field can be used to distinguish between different types of motion areas. As mentioned in Sand and Teller (2008b), the divergence of a flow field is negative for occluded areas, positive for disoccluded, and near zero for the matched areas.

Let us consider 3 consecutives gray level frames $I_{-1}, I_0, I_1 : \Omega \to \mathbb{R}$ in order to compute the optical flow between I_0, I_1. Let $\chi : \Omega \to [0, 1]$ be the function modeling the occlusion mask, so that $\chi = 1$ identifies the occluded pixels, i.e. pixels that are visible in I_0 but not in I_1. Our model is based on the assumptions

(i) Pixels that are not visible in frame I_1 are visible in the previous frame of I_0. Let $I_{-1} : \Omega \to \mathbb{R}$ be that frame.

(ii) Motions of the occluded background area are not fast.

Thus, if $\chi(x) = 0$, then we compare $I_0(x)$ and $I_1(x + u(x))$. If $\chi(x) = 1$, we compare $I_0(x)$ and $I_{-1}(x - u(x))$. On the other hand, the occluded region given by $\chi = 1$ should be correlated with the region where $\text{div}(u)$ is negative. Thus we propose to compute the optical flow and the occlusion mask by minimizing the energy

$$E(u, \chi) = E_d(u, \chi) + E_r(u, \chi) + \frac{\eta}{2} \int_\Omega \chi \, |u|^2 \, dx + \beta \int_\Omega \chi \text{div}(u) dx, \quad (13.2)$$

where

$$E_d(u,\chi) = \lambda \int_\Omega ((1-\chi(x))\,|I_0(x) - I_1(x+u(x))|+\chi(x)\,|I_0(x) - I_{-1}(x-u(x))|)dx,$$

and

$$E_r(u,\chi) = \int_\Omega g(x)(|\nabla u_1| + |\nabla u_2| + |\nabla \chi|)dx, \qquad (13.3)$$

with $\eta \geq 0$, $\beta > 0$ and $g(x) = \frac{1}{1+\gamma|\nabla I_0(x)|}$, $x \in \Omega$, $\gamma > 0$ (g can be also choosen to be $g(\mathbf{x}) = 1$ or $g(x) = \frac{1}{1+\gamma|\nabla \tilde{I}_0(x)|}$, where $\tilde{I}_0$ is a smoothed version of I_0). We have included in (13.2) a term penalizing large displacements where $\chi = 1$ (with $\eta > 0$ but small relative to λ). This is motivated by two observations. On one hand, we are assuming that the occluded background area is moving slower than the occluding foreground. On the other, since images have usually self-similarities, a pixel may have several possibilities to match. Thus, with $\eta > 0$ we try to avoid false matchings due to self-similarities in the image by encouraging to choose the smallest displacement (in practice we take $\eta = 0.01$).

In order to cope with the nonlinearities in the energy terms, we propose to use an auxiliary variable v representing the optical flow and we penalize its deviation from u. That is, we introduce a term $\frac{1}{2\theta} \int_\Omega |u(x) - v(x)|^2$ and, finally, the following energy is proposed

$$E_\theta(u,v,\chi) = E_d(v,\chi) + E_r(u,\chi)+$$
$$\frac{\eta}{2}\int_\Omega \chi\,|v|^2\,dx + \beta\int_\Omega \chi div(u)dx + \frac{1}{2\theta}\int_\Omega |u-v|^2\,dx, \quad (13.4)$$

depending on three variables, u, v and χ. This decoupled energy can be minimized by alternatively fixing two variables and minimizing with respect to the third one.

To minimize (13.4) with respect to v, we linearize each expression $|I_0(x) - I_i(x+\epsilon_i u(x))|$, $i = -1, 1$ ($\epsilon_{-1} = -1, \epsilon_1 = 1$), around a given vector field u_0 (close approximation of u) and define the residual

$$\rho_i(v) = I_0(x) - I_i(x+\epsilon_i u_0(x)) - \epsilon_i \nabla I_i(x+\epsilon u_0(x))(v(x) - u_0(x)). \quad (13.5)$$

This approximation procedure is applied by alternatively minimizing the energy model:

$$\tilde{E}_\theta(u, v, \chi) = \tilde{E}_d(v, \chi) + E_r(u, \chi) +$$
$$\frac{\eta}{2} \int_\Omega \chi |v|^2 \, dx + \beta \int_\Omega \chi \, div(u) dx + \frac{1}{2\theta} \int_\Omega |u - v|^2 \, dx, \quad (13.6)$$

with

$$\tilde{E}_d(v, \chi) = \lambda \int_\Omega ((1 - \chi) |\rho_1(v)| + \chi |\rho_{-1}(v)|) dx,$$

where η, λ, β, θ are positive real constants.

The energy $\tilde{E}_\theta(u, v, \chi)$ is minimized using an alternated minimization strategy. That is, minimizing the energy with respect to one variable while keeping the other two fixed. This optimization procedure is given by Propositions 13.1, 13.2 and 13.3 below. After iteration of these steps, we proceed to redefine $\rho_i(v)$ (see Algorithm 13.8 in Section 13.3).

The minimization of $\widetilde{E}_\theta$ with respect to $\mathbf{u}$ is done using Chambolle's algorithm Chambolle (2004).

Proposition 13.1. *The minimum of $\tilde{E}_\theta$ with respect to $\mathbf{u} = (u_1, u_2)$ is given by*

$$u_i = v_i + \theta \mathrm{div}(g\xi_i) + \theta\beta \frac{\partial \chi}{\partial x_i}, \quad i = 1, 2, \quad (13.7)$$

ξ_1 and ξ_2 are computed using the following iterative scheme

$$\xi_i^{k+1} = \frac{\xi_i^k + \frac{\tau_u}{\theta} g \nabla(v_i + \theta \mathrm{div}(g\xi_i^k) + \theta\beta \frac{\partial \chi}{\partial x_i})}{1 + \frac{\tau_u}{\theta} |g \nabla(v_i + \theta \mathrm{div}(g\xi_i^k) + \theta\beta \frac{\partial \chi}{\partial x_i})|}, \quad k = 0, 1, 2, \dots \quad (13.8)$$

where $\xi_i^0 = 0$ and $\tau_u \leq 1/8$.

As in Zach et al. (2007), we can solve explicitly with respect to v:

Proposition 13.2. *Assume that $\chi : \Omega \to \{0, 1\}$. The minimum of $\tilde{E}_\theta$ with respect to $\mathbf{v} = (v_1, v_2)$ is*

$$\mathbf{v} = \begin{cases} \eta_i \mathbf{u} - \mu_i \epsilon_i \nabla I_i(\mathbf{x}^*) & \text{if} \quad \Lambda_i(\mathbf{u}) > \mu_i |\nabla I_i(\mathbf{x}^*)|^2 \\ \eta_i \mathbf{u} + \mu_i \epsilon_i \nabla I_i(\mathbf{x}^*) & \text{if} \quad \Lambda_i(\mathbf{u}) < -\mu_i |\nabla I_i(\mathbf{x}^*)|^2 \\ \mathbf{u} - \epsilon_i \rho_i(\mathbf{u}) \frac{\nabla I_i(\mathbf{x}^*)}{|\nabla I_i(\mathbf{x}^*)|^2} & \text{if} \quad |\Lambda_i(\mathbf{u})| \leq \mu_i |\nabla I_i(\mathbf{x}^*)|^2, \end{cases} \quad (13.9)$$

where $\mathbf{x}^* = \mathbf{x} + \epsilon_i \mathbf{u}_0$ *and where* $i = 1$ *and* $\epsilon_1 = 1$, $\eta_1 = 1$, $\mu_1 = \lambda\theta$, $\Lambda_1(\mathbf{u}) = \rho_1(\mathbf{u})$ *when* $\chi = 0$, *and* $i = -1$, $\epsilon_{-1} = -1$, $\eta_{-1} = \frac{1}{1+\eta\theta}$, $\mu_{-1} = \frac{\lambda\theta}{1+\eta\theta}$, $\Lambda_{-1}(\mathbf{u}) = \rho_{-1}(\mathbf{u}) + \frac{\eta\theta}{1+\eta\theta} \mathbf{u} \cdot \nabla I_{-1}(\mathbf{x} + \epsilon_i \mathbf{u}_0)$ *when* $\chi = 1$.

Notice that we omitted the arguments $\mathbf{x}$ *in* $\mathbf{u}, \mathbf{u}_0$.

Having computed $\mathbf{v}$, let $F = \lambda(|\rho_{-1}(\mathbf{v})| - |\rho_1(\mathbf{v})|)$ and $G = \frac{\eta}{2}|\mathbf{v}|^2$. As a consequence of Chambolle and Pock (2011), we have

Proposition 13.3. *Let* $0 < \tau_\psi \tau_\chi < 1/8$. *Given* $\mathbf{u}, \mathbf{v}$, *the minimum* $\bar{\chi}$ *of* $\tilde{E}_\theta$ *with respect to* χ *can be obtained by the following primal-dual algorithm*

$$\begin{aligned}
\psi^{n+1} &= P_B(\psi^n + \tau_\psi \, g \, \nabla\chi^n) \\
\chi^{n+1} &= P_{[0,1]}\left(\chi^n + \tau_\chi \left(\operatorname{div}(g\psi^{n+1}) - \beta \operatorname{div}\mathbf{u} - F - G\right)\right),
\end{aligned} \tag{13.10}$$

where $P_B(\psi)$ *denotes the projection of* ψ *on the unit ball of* $\mathbb{R}^2$ *and* $P_{[0,1]}(r) = \max(\min(r,1),0)$, $r \in \mathbb{R}$.

Notice that, by the co-area formula, the level sets of χ are also minimizers of $\tilde{E}_\theta$ ($\mathbf{u}, \mathbf{v}$ being fixed). Thus, before going to next minimization of $\tilde{E}_\theta$ with respect to $\mathbf{u}$, we redefine $\chi(\mathbf{x}) = T_\delta(\bar{\chi}(\mathbf{x}))$, where $T_\delta(r) = 1$ (resp. 0) if $r \geq \delta$ (resp. $< \delta$). A different relaxation that also produces good results is obtained by replacing $1 - \chi$ and χ by $(1-\chi)^2$ and χ^2, respectively, in all terms of (13.6) but E_r.

13.2 A Model Considering Occlusions, Illumination Changes and Large Displacement

13.2.1 An optical flow and occlusions model for color video

Let us start by extending the occlusion and optical flow model presented above to the case of color image sequences. The energy model will use the color constancy assumption. We consider 3 consecutives color frames $I_{-1}, I_0, I_1 : \Omega \to \mathbb{R}^3$, which we assume to have values in the RBG color space, hence each frame I_i has three color components I_i^1, I_i^2, I_i^3, associated to the red, green and blue channels, respectively. Thus, the model to jointly compute the optical flow and occlusions becomes

$$E^c(u, \chi) = E_d^c(u, \chi) + E_r(u, \chi) + \frac{\eta}{2} \int_\Omega \chi |u|^2 \, dx + \beta \int_\Omega \chi \operatorname{div}(u) \, dx, \tag{13.11}$$

where

$$E_d^c(u, \chi) = \lambda \sum_{k=1}^{3} \int_\Omega ((1-\chi)|I_0^k(x) - I_1^k(x+u(x))| + \chi|I_0^k(x) - I_{-1}^k(x-u(x))|) \, dx,$$

(13.12)

In order to minimize (13.11), we relax it and introduce three auxiliary variables v_1, v_2, v_3 representing the flow and used to decouple the nonlinear terms, where v_1, v_2, v_3 correspond the red, green and blue components, respectively. As before, we penalize the difference between the optical flow u and each of the auxiliary variables v_1, v_2, v_3. To simplify the presentation, we use the same notation as in Xu. et al. (2010) and concatenate the vectors in $\tilde{v}(x) = \{v_1(x), v_2(x), v_3(x)\}^T \in \mathbb{R}^{3 \times 2}$, $I(x) = \{I^1(x), I^2(x), I^3(x)\}^T \in \mathbb{R}^3$, $\tilde{u}(x) = \{u(x), u(x), u(x)\}^T \in \mathbb{R}^{3 \times 2}$. In this context, let us remark that the vector $I(x+\tilde{v})$ denotes $I(x+\tilde{v}) = \{I^1(x+v_1), I^2(x+v_2), I^3(x+v_3)\}^T$.

Thus, to compute the occlusions and the optical flow between I_0, I_1, we propose to minimize the following energy:

$$E^c(u, \chi, \tilde{v}) = E_d^c(\tilde{v}, \chi) + E_r(u, \chi) +$$
$$\frac{\eta}{2} \int_\Omega \chi \, |\tilde{v}|^2 \, dx + \beta \int_\Omega \chi div(u) dx + \frac{1}{2\theta} \int_\Omega |\tilde{u} - \tilde{v}|^2 \, dx, \quad (13.13)$$

where $|\tilde{v}|^2$ stands for $\sum_{k=1}^{3} |v_k|^2$ and

$$E_d^c(\tilde{v}, \chi) =$$
$$\lambda \int_\Omega (1 - \chi) \sum_{k=1}^{3} \left| \rho_1^k(v_k) \right| dx + \lambda \int_\Omega \chi \sum_{k=1}^{3} \left| \rho_{-1}^k(v_k) \right| dx, \quad (13.14)$$

and ρ_i^k is the linearized version of $I_0^k(x) - I_i^k(\mathbf{x} + \epsilon_i \mathbf{v}_k)$ around an approximation u_0 of u, with $i = -1, 1$ and $\epsilon_{-1} = -1$ and $\epsilon_1 = 1$, and $k = 1, 2, 3$ (corresponding to each color channel). The linearization procedure is applied to each $\rho_i^k(\mathbf{x})$.

As in previous Section 13.1, $\lambda, \beta > 0$ and $\eta \geq 0$. When $\eta > 0$ the model penalizes large displacements in places where $\chi = 1$ trying to avoide false

matchings due to self-similarities in the image. In practice, η is choosen to be small relative to λ. Finally, the term $E_r(u, \chi)$ is given by

$$E_r(u, \chi) = \int_\Omega g(x)(|\nabla u_1(x)| + |\nabla u_2(x)| + |\nabla \chi(x)|)\, dx, \qquad (13.15)$$

where $g(x) = \frac{1}{1+\gamma|\nabla I_0^g(x)|}$, $x \in \Omega$, $\gamma > 0$, and I_0^g is the gray level version of the frame at $t = 0$.

Minimizing the Energy

As in Section 13.1, we minimize $\widetilde{E}^c$ in (13.13) by alternating among the minimization with respect to each variable while keeping the other two fixed. In particular, the minimization of $\widetilde{E}^c$ with respect to u, v^k and χ is described in the following Propositions.

Proposition 13.4. *The minimum of $\widetilde{E}_\alpha^c$ with respect to $\mathbf{u} = (u_1, u_2)$ is given by*

$$u_i = \frac{1}{3}\sum_{k=1}^{3} v_k^i + \theta \operatorname{div}(g\xi_i) + \theta\beta\frac{\partial\chi}{\partial x_i}, \quad i = 1, 2, \qquad (13.16)$$

ξ_1 and ξ_2 are computed using the following iterative scheme

$$\xi_i^{t+1} = \frac{\xi_i^t + \frac{\tau_u}{\theta}g\nabla(\frac{1}{3}\sum_{k=1}^{3} v_k^i + \theta\operatorname{div}(g\xi_i^t) + \theta\beta\frac{\partial\chi}{\partial x_i})}{1 + \frac{\tau_u}{\theta}|g\nabla(\frac{1}{3}\sum_{k=1}^{3} v_k^i + \theta\operatorname{div}(g\xi_i^t) + \theta\beta\frac{\partial\chi}{\partial x_i})|}, \quad t = 0, 1, 2, \ldots$$
$$(13.17)$$

where $\xi_i^0 = 0$ and $\tau_u \leq 1/8$.

Proposition 13.5. *Assume that $\chi : \Omega \to \{0, 1\}$. The minimum of $\tilde{E}_\theta$ with respect to $\mathbf{v_k} = (v_k^1, v_k^2)$ is*

$$\mathbf{v_k} = \begin{cases} \eta_i\mathbf{u} - \mu_i\epsilon_i\alpha(x)\nabla I_i^k(\mathbf{x}^*) & \text{if} \quad \Lambda_i^k(\mathbf{u}) > \mu_i\alpha(x)|\nabla I_i^k(\mathbf{x}^*)|^2 \\ \eta_i\mathbf{u} + \mu_i\epsilon_i\alpha(x)\nabla I_i^k(\mathbf{x}^*) & \text{if} \quad \Lambda_i^k(\mathbf{u}) < -\mu_i\alpha(x)|\nabla I_i^k(\mathbf{x}^*)|^2 \\ \mathbf{u} - \epsilon_i\rho_i^k(\mathbf{u})\frac{\nabla I_i^k(\mathbf{x}^*)}{|\nabla I_i^k(\mathbf{x}^*)|^2} & \text{if} \quad |\Lambda_i^k(\mathbf{u})| \leq \mu_i\alpha(x)|\nabla I_i^k(\mathbf{x}^*)|^2, \end{cases}$$
$$(13.18)$$

when $i = 1$ and $\epsilon_1 = 1$, $\eta_1 = 1$, $\mu_1 = \lambda\theta$, $\Lambda_1^k(\mathbf{u}) = \rho_1^k(\mathbf{u})$ when $\chi = 0$, and $i = -1$, $\epsilon_{-1} = -1$, $\eta_{-1} = \frac{1}{1+\eta\theta}$, $\mu_{-1} = \frac{\lambda\theta}{1+\eta\theta}$, $\Lambda_{-1}(\mathbf{u}) = \rho_{-1}^k(\mathbf{u}) + \frac{\eta\theta}{1+\eta\theta}\mathbf{u} \cdot \nabla I_{-1}^k(\mathbf{x} + \epsilon_i\mathbf{u}_0)$ when $\chi = 1$. Additionally we create $\mathbf{x}^* = \mathbf{x} + \epsilon_i\mathbf{u}_0$.

The arguments $\mathbf{x}$ in $\mathbf{u}, \mathbf{u}_0$ are omitted.

Once all v_k are computed, we define

$$F = -\lambda \sum_{k=1}^{3} \left| \rho_1^k \right| + \frac{1}{3} \sum_{k=1}^{3} (v_k - u)^2, \qquad (13.19)$$

and

$$G = -\lambda \sum_{k=1}^{3} \left| \rho_{-1}^k \right| + \frac{\eta}{3} \sum_{k=1}^{3} (v_k)^2. \qquad (13.20)$$

Proposition 13.6. *Let $0 < \tau_\psi \tau_\chi < 1/8$. Given $\mathbf{u}, \mathbf{v}$, the minimum $\bar{\chi}$ of $\tilde{E}_\theta$ with respect to χ can be obtained by the following primal-dual algorithm*

$$\begin{aligned} \psi^{n+1} &= P_B(\psi^n + \tau_\psi \; g \; \nabla \chi^n) \\ \chi^{n+1} &= P_{[0,1]} \left(\chi^n + \tau_\chi \left(\mathrm{div}(g\psi^{n+1}) - \beta \; \mathrm{div}\mathbf{u} - F - G \right) \right), \end{aligned} \qquad (13.21)$$

where $P_B(\psi)$ denotes the projection of ψ on the unit ball of $\mathbb{R}^2$ and $P_{[0,1]}(r) = \max(\min(r,1),0)$, $r \in \mathbb{R}$.

13.2.2   Adding robustness to illumination changes

The color constancy assumption is frequently violated due to illumination changes, shadows or specular reflections. A combination of the color constancy assumption and the gradient constancy assumption in the data term seems to be a valuable approach to alleviate this problem Xu. et al. (2010). The gradient constancy assumption was already used in Brox et al. (2004b) to obtain robustness against additive illumination changes. A good way to introduce a combination of color and gradient is to use an adaptive weight map $\alpha : \Omega \to [0,1]$ that allows to adaptively balance the contribution of color and gradient constraints at each point in the image domain. We propose the following model:

$$\begin{aligned} E_\alpha^c(u,\chi) = E_{d,\alpha}^c(u,\chi) + E_r(u,\chi) + \\ \frac{\eta}{2} \int_\Omega \chi |u|^2 dx + \beta \int_\Omega \chi div(u) dx, \quad (13.22) \end{aligned}$$

where $E_{d,\alpha}^c(u,\chi)$ can be written as

$$E_{d,\alpha}^c(u,\chi) = \int_\Omega \left(\alpha(x) D_{I,\chi}(u,\chi,x) + (1-\alpha(x)) D_{\nabla I,\chi}(u,\chi,x) \right) dx \quad (13.23)$$

and $D_{I,\chi}(u, \chi, x)$ and $D_{\nabla I,\chi}(u, \chi, x)$ are pointwise data costs based on the comparison of color and gradients of the image, respectively. Roughly speaking, $D_{I,\chi}$ contains the comparison $\|I_0(x) - I_1(x + u)\|$ and $D_{\nabla I,\chi}$ the comparison $\tau\|\nabla I_0(x) - \nabla I_1(x+u)\|$, with $\tau > 0$. We will detail below how these terms are defined depending on χ. Therefore, the data term in (13.22) is made of two parts and $\alpha(x)$ provides an ad balance term between intensity and gradient constraints. For instance, when $\alpha(x) = 0$, the gradient constraint is favoured.

To provide a compact presentation, we define a new vector $\tilde{I}$ that includes the gradient, $\tilde{I} = \{I_i^1, I_i^2, I_i^3, \partial_x I_i, \partial_y I_i\}$. Where $\partial_x I_i$ and $\partial_y I_i$ represents partial derivatives with respect to x and y, respectively, of the gray level frame I_i, for $i = 1, -1$. Given $x \in \Omega$, from the weight map value $\alpha(x)$, we define a new vector $\tilde{\alpha}(x)$ to represent the corresponding weights for color and gradient constancy assumptions, namely, $\tilde{\alpha} = \{\alpha, \alpha, \alpha, (1 - \alpha)\tau, (1 - \alpha)\tau\}$, where $\tau > 0$ is a constant value and we have omitted the point x and written $\alpha = \alpha(x)$. Next section presents the definition of α used in our model. For a more compact notation, let us denote it as $\tilde{\alpha}(x) = \{\alpha^1, \alpha^2, \alpha^3, \alpha^4, \alpha^5\}$, where $\alpha^1 = \alpha^2 = \alpha^3 = \alpha$ and $\alpha^4 = \alpha^5 = (1 - \alpha)\tau$.

Then, we propose to add auxiliary variables $v_k, k = 1, \ldots 5$, penalize its deviation from the optical flow u and minimize the energy

$$\widetilde{E}_\alpha^c(u, \chi, \tilde{v}) = E_{d,\alpha}^c(\tilde{v}, \chi) + E_r(u, \chi) +$$
$$\frac{\eta}{2} \int_\Omega \chi \sum_{i=1}^5 |v_k|^2 \, dx + \frac{1}{2\theta} \int_\Omega \sum_{k=1}^5 |u - v_k|^2 \, dx + \beta \int_\Omega \chi div(u) dx. \quad (13.24)$$

where $\theta > 0$ and the new term $E_{d,\alpha}^c$ that considers the color constancy assumption, the gradient constancy assumption and the weight map $\tilde{\alpha}$ is defined as

$$E_{d,\alpha}^c(\tilde{v}, \chi) = \lambda \int_\Omega \left[(1 - \chi) \sum_{k=1}^3 \alpha^k(x) \left| \rho_1^k(v_k) \right| dx + \chi \sum_{k=1}^3 \alpha^k(x) \left| \rho_{-1}(v_k) \right| \right] dx$$
$$+ \lambda \int_\Omega \left[(1 - \chi) \sum_{k=4}^5 \alpha^k(x) \left| \rho_1^k(v_k) \right| + \chi \sum_{k=4}^5 \alpha^k(x) \left| \rho_{-1}^k(v_k) \right| \right] dx. \quad (13.25)$$

For $k = 1, 2, 3$, $\rho_i^k(v_k)$ is a linearized version of $I_0^k(x) - I_i^k(x + \epsilon_i u)$ around an approximation u_0 of u. For $k = 4, 5$, $\rho_i^k(v_k)$ is a linearized version of $\partial_x I_0(x) - \partial_x I_i(x + \epsilon_i u_0)$ and $\partial_y I_0(x) - \partial_y I_i(x + \epsilon_i u_0)$, respectively, given by:

$$\rho_i^4(v_4) = \partial_x I_0 - \partial_x I_i(\mathbf{x}^*) + \epsilon_i \nabla \left(\partial_x I_i(x^*) \right) (v_4(x) - u_0(x))$$

and

$$\rho_i^5(v_5) = \partial_y I_0 - \partial_y I_i(\mathbf{x}^*) + \epsilon_i \nabla \left(\partial_y I_i \right) (x^*)(v_5(x) - u_0(x)).$$

Where $x^* = x + \epsilon_i u_0(x)$ with $i = -1, 1$ ($\epsilon_{-1} = -1, \epsilon_1 = 1$).

An adaptative weight map $\alpha(x)$ to balance color and gradient constraints

The proposed model (13.24) depends on the definition of $\alpha(x)$, the function which selectively combines the color and gradient information. In Xu. et al. (2010) the computation of $\alpha(x)$ is explicit and depends on the difference between two terms $D_I(u, x)$ and $D_{\nabla I}(u, x)$. The term $D_I(u, x)$ represents the color constancy assumption and $D_{\nabla I}(u, x)$ represents the gradient constancy assumption. In particular,

- The expression for the color constancy assumption, $D_I(u, x)$ is given by:

$$D_I(u, x) = \sum_{k=1}^3 \left| I_0^k(x) - I_1^k(x + u) \right| \tag{13.26}$$

  where I_0^k and I_1^k are the RGB components of two consecutive color images.

- The term $D_{\nabla I}(u, x)$ is computed as:

$$D_{\nabla I}(u, x) = \tau \left(|\partial_x I_0(x) - I_1(x + u)\partial_x| + |\partial_y I_0(x) - \partial_y I_1(x + u)| \right), \tag{13.27}$$

  where I_0 and I_1 are the gray level version of the images in $D_I(u, x)$.

Then, the weight map $\alpha(x)$ is defined in Xu. et al. (2010) as

$$\alpha(x) = \frac{1}{1 + e^{\tilde{\beta}(D_I(u,x) - D_{\nabla I}(u,x))}}, \tag{13.28}$$

where $\tilde{\beta}$ is a positive constant. Let us comment about the behaviour of (13.28). If the term $D_I(u, x) \gg D_{\nabla I}(u, x)$, the difference $D_I(u, x) - D_{\nabla I}(u, x)$ will be positive and the exponential value $e^{\tilde{\beta}(D_I(u,x) - D_{\nabla I}(u,x))}$ will be large. Then, $\alpha(x)$ will be a small value, say near 0, and the data term will have more confidence on the gradient constancy assumption. On the other hand, if $D_{\nabla I}(u, x) \gg D_I(u, x)$, the difference $D_I(u, x) - D_{\nabla I}(u, x)$ will be negative and the exponential value $e^{\tilde{\beta}(D_I(u,x) - D_{\nabla I}(u,x))}$ will be very small. Then, $\alpha(x)$ will produce a value near to 1. In other words, the data term will be more confident on the color constancy assumption.

Our proposal extends definitions (13.26) and (13.27) for D_I and $D_{\nabla I}$, respectively, to consider our occlusion layer χ. We define the term $D_{I,\chi}$ as

$$D_{I,\chi}(u, \chi, x) = \sum_{k=1}^{3}(1 - \chi)\left|I_0^k(x) - I_1^k(x + u)\right| + \chi\left|I_0^k(x) - I_{-1}^k(x - u)\right|. \tag{13.29}$$

We can rewrite (13.29) using a linearized version of the difference $I_0^k(x) - I_i^k(x + \epsilon_i v(x))$ (where $k = 1, 2, 3$ indicates the color component of the image, and $i = 1, -1$),

$$D_{I,\chi}(u, \chi, x) = \sum_{k=1}^{3}(1 - \chi)\left|\rho_1^k\right| + \chi\left|\rho_{-1}^k\right|. \tag{13.30}$$

Finally, $D_{\nabla I,\chi}(u, \chi, x)$ extends the gradient constancy assumption term in order to consider the occlusion mask by

$$D_{\nabla I,\chi}(u, \chi, x) = (1 - \chi)\tau\sum_{k=4}^{5}(\left|I_0^k(x) - I_1^k(x + u)\right| +$$
$$\chi\tau\sum_{k=4}^{5}(\left|I_0^k(x) - I_{-1}^k(x - u)\right|, \quad \text{(13.31)}$$

where $\tau > 0$ and I_i^k, for $k = 4$ and $k = 5$ denote the partial derivatives with respect to x and y, respectively, computed on the gray level frame I_i, for $i = -1, 0, 1$. Finally,

$$\alpha(x) = \frac{1}{1 + e^{\tilde{\beta}(D_{I,\chi}(u,\chi,x) - D_{\nabla I,\chi}(u,\chi,x))}}, \tag{13.32}$$

with $\tilde{\beta} > 0$, and the proposed energy model for robust computation of optical flow and occlusions is given by (13.24) presented at the beginning of this section, that is,

$$
\widetilde{E}_\alpha^c(u, \chi, \tilde{v}) = E_{d,\alpha}^c(\tilde{v}, \chi) + E_r(u, \chi)+
$$

$$
\frac{\eta}{2} \int_\Omega \chi \sum_{i=1}^5 |v_k|^2 \, dx + \frac{1}{2\theta} \int_\Omega \sum_{k=1}^5 |u - v_k|^2 \, dx + \beta \int_\Omega \chi \, div(u) dx. \quad (13.33)
$$

Minimizing the Energy

As in Section 13.1, we minimize $\widetilde{E}_\alpha^c$ in (13.33) by alternating among the minimization with respect to each variable while keeping the other two fixed. In particular, the minimization of $\widetilde{E}_\alpha^c$ with respect to u, v^k and χ is described in the following Propositions.

Proposition 13.7. *The minimum of $\widetilde{E}_\alpha^c$ with respect to $\mathbf{u} = (u_1, u_2)$ is given by*

$$
u_i = \frac{1}{5} \sum_{k=1}^5 v_k^i + \theta \mathrm{div}(g\xi_i) + \theta\beta \frac{\partial \chi}{\partial x_i}, \quad i = 1, 2, \quad (13.34)
$$

ξ_1 and ξ_2 are computed using the following iterative scheme

$$
\xi_i^{t+1} = \frac{\xi_i^t + \frac{\tau_u}{\theta} g \nabla(\frac{1}{5} \sum_{k=1}^5 v_k^i + \theta \mathrm{div}(g\xi_i^t) + \theta\beta \frac{\partial \chi}{\partial x_i})}{1 + \frac{\tau_u}{\theta} |g \nabla(\frac{1}{5} \sum_{k=1}^5 v_k^i + \theta \mathrm{div}(g\xi_i^t) + \theta\beta \frac{\partial \chi}{\partial x_i})|}, \quad t = 0, 1, 2, ...
$$

$$
(13.35)
$$

where $\xi_i^0 = 0$ and $\tau_u \leq 1/8$.

Proposition 13.8. *Assume that $\chi : \Omega \to \{0, 1\}$. The minimum of $\tilde{E}_\theta$ with respect to $\mathbf{v_k} = (v_k^1, v_k^2)$ is*

$$
\mathbf{v_k} = \begin{cases}
\eta_i \mathbf{u} - \mu_i \epsilon_i \alpha(x) \nabla I_i^k(\mathbf{x}^*) & \text{if} \quad \Lambda_i^k(\mathbf{u}) > \mu_i \alpha(x)|\nabla I_i^k(\mathbf{x}^*)|^2 \\
\eta_i \mathbf{u} + \mu_i \epsilon_i \alpha(x) \nabla I_i^k(\mathbf{x}^*) & \text{if} \quad \Lambda_i^k(\mathbf{u}) < -\mu_i \alpha(x)|\nabla I_i^k(\mathbf{x}^*)|^2 \\
\mathbf{u} - \epsilon_i \rho_i^k(\mathbf{u}) \frac{\nabla I_i^k(\mathbf{x}^*)}{|\nabla I_i^k(\mathbf{x}^*)|^2} & \text{if} \quad |\Lambda_i^k(\mathbf{u})| \leq \mu_i \alpha(x)|\nabla I_i^k(\mathbf{x}^*)|^2,
\end{cases}
$$

$$
(13.36)
$$

when $i = 1$ and $\epsilon_1 = 1$, $\eta_1 = 1$, $\mu_1 = \lambda\theta$, $\Lambda_1^k(\mathbf{u}) = \rho_1^k(\mathbf{u})$ when $\chi = 0$, and $i = -1$, $\epsilon_{-1} = -1$, $\eta_{-1} = \frac{1}{1+\eta\theta}$, $\mu_{-1} = \frac{\lambda\theta}{1+\eta\theta}$, $\Lambda_{-1}(\mathbf{u}) = \rho_{-1}^k(\mathbf{u}) + \frac{\eta\theta}{1+\eta\theta}\mathbf{u} \cdot \nabla I_{-1}^k(\mathbf{x} + \epsilon_i\mathbf{u}_0)$ when $\chi = 1$. Additionally we create $\mathbf{x}^* = \mathbf{x} + \epsilon_i\mathbf{u}_0$.

The arguments $\mathbf{x}$ in $\mathbf{u}, \mathbf{u}_0$ are omitted.

Figure 13.1: Images of the sequence Ambush_7. First row: frames 1, 7, 11. Second row: frames 32, 41, 49.

Once all v_k are computed, we define

$$F = \lambda \left(-\alpha(x) \sum_{k=1}^{3} \left| \rho_1^k \right| - (1 - \alpha(x)) \sum_{k=4}^{5} |\rho_1(v_k)| \right) + \frac{1}{5} \left(\sum_{k=1}^{5} (v_k - u)^2 \right),$$
(13.37)

and

$$G = \lambda \left(-\alpha(x) \sum_{k=1}^{3} \left| \rho_{-1}^k \right| - (1 - \alpha(x)) \sum_{k=4}^{5} |\rho_{-1}(v_k)| \right) + \frac{\eta}{5} \left(\sum_{k=1}^{5} (v_k)^2 \right).$$
(13.38)

Proposition 13.9. *Let $0 < \tau_\psi \tau_\chi < 1/8$. Given $\mathbf{u}, \mathbf{v}$, the minimum $\bar{\chi}$ of $\tilde{E}_\theta$ with respect to χ can be obtained by the following primal-dual algorithm*

$$\begin{aligned} \psi^{n+1} &= P_B(\psi^n + \tau_\psi \ g \ \nabla \chi^n) \\ \chi^{n+1} &= P_{[0,1]} \left(\chi^n + \tau_\chi \left(\mathrm{div}(g\psi^{n+1}) - \beta \ \mathrm{div}\mathbf{u} - F - G \right) \right), \end{aligned}$$
(13.39)

where $P_B(\psi)$ denotes the projection of ψ on the unit ball of $\mathbb{R}^2$ and $P_{[0,1]}(r) = \max(\min(r, 1), 0)$, $r \in \mathbb{R}$.

Let us assume that we have already computed an optical flow u minimizing the functional (13.33). We refer to it as u_{rc} (with rc standing for optical flow for color images and robust to brightness changes).

As a proof of concept we present here a first validation of the performance of this proposal for optical flow estimation. We present in Figure 13.1 some frames of the sequence Ambush_7 of MPI Sintel database Butler et al. (2012). Sintel is a database of realistic video sequences from an open source animated short film. It contains shadows, reflections and illumination, as

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)
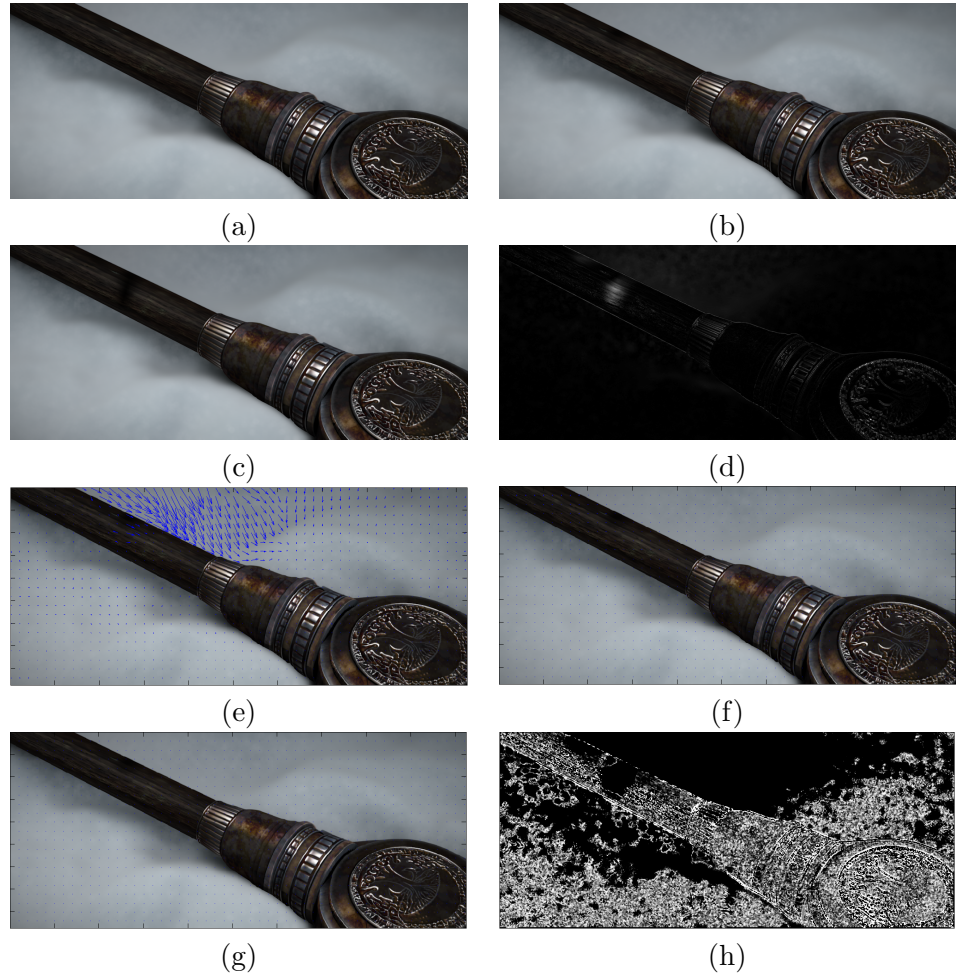
Figure 13.2: Effect of the adaptative $\alpha(x)$. From left to right and top to bottom, (a), (b), (c) show three consecutive frames where there are illumination changes and shadows. (d) shows the absolute color difference between (b) and (c). Then, (e), (f) and (g) show the optical flow obtained using $\alpha(x) = 1.0$, $\alpha(x) = 0.0$, and adaptative $\alpha(x)$, respectively. Finally, (h) displays this adaptative map $\alpha(x)$.

well as other challenges. We will give more details of MPI Sintel in chapter 14. In the video sequence shown in Figure 13.1, a girl takes a lance that lays on the snow. In the five first frames of the video Ambush_7, the hand of the girl approaches the lance. To empirically show the power of our proposal, in Figure 13.2 we focus on three frames of the first part of this sequence. In frames 2, 3 and 4, the hand is not in the scene (see Figure 13.2) but the shadow of the hand appears in the scene and does move. In other words, in those three frames there is no real movement in the scene only a shadow that moves. In order to illustrate the effect of the adaptative weight $\alpha(x)$ in the data term, we have computed the optical flow for these three frames by minimizing (13.33) but with three differents options for $\alpha(x)$, namely, $\alpha(x) = 1.0$, $\alpha(x) = 0.0$ and our adaptive $\alpha(x)$. In Figure 13.2(a), (b) and (c), we first display the three consecutive frames used. In order to display the changes in intensities and shadows, Figure 13.2(d) shows the magnitude of the color difference $\|I_0(x) - I_1(x)\|$ between frames (b) (say I_0) and (c) (say I_1). For visualization purposes, the absolute differences values were scaled between 0 and 255. Figure 13.2(e), (f) and (g) display the estimated optical flow obtained when considering $\alpha(x) = 1.0$, $\alpha(x) = 0.0$ and adaptive $\alpha(x)$, respectively. The optical flow is represented with arrows which length is proportional to the optical flow modulus. Let us notice that in the region where shadows moves in the video sequence the optical flow is incorrectly computed when $\alpha(x) = 1.0$ (only intensities) but the flow is correctly computed using gradients. Finally, Figure 13.2(h) displays the values of our adaptative $\alpha(x)$. Let us remark that in regions where shadows moves, we obtain $\alpha(x) = 0.0$, as can be seen on the snow and on the lance handle. On other hand, let us notice that the average of the $\alpha(x)$ values in the interval $0.1 < \alpha(x) < 1$ is 0.5332. A value of $\alpha(x) = 0.5$ means that the terms $D_{I,\chi}$ and $D_{\nabla I,\chi}$ have a similar role in that regions and thus both might be used.

13.2.3 Large displacement method

A drawback of the variational model described above is its incapability to handle large displacements of small objects that move differently from its neighbors. Indeed, in order to estimate large displacements using variational approaches, the optical flow computation is usually embedded in a multi-scale optimization strategy. As commented previously, the idea behind a multi-scale approach is to create a coarse-to-fine structure that enables the estimation of the flow field at coarser scales and then to refine the solution at finer scales Mémin and Pérez (1998); Brox et al. (2004b);

Meinhardt-Llopis et al. (2013). But in this case, the method is unable to
estimate large displacements which are larger than the size of the object
that moves, specially if the small object moves differently from its neigh-
bours. As reviewed in Chapter 12, in recent years research has been done
for handling large displacements Xu. et al. (2010); Steinbruecker and Pock
(2009); Weinzaepfel et al. (2013); Timofte and Van Gool (2015). For in-
stance, the work in Steinbruecker and Pock (2009) is based on exhaustive
search performing block matching. In Stoll et al. (2012) the estimation of
optical flow is improved by the integration of supplementary matchings.
These supplementary matchings are used only at specific locations where
they are really needed and finally integrated into the variational model.

In this section we propose a new optical flow model for large displacement
and occlusion estimation. To this goal, we improve our robust color model
by combining it with exhaustive search. By using a pointwise confidence
measure, our strategy directly integrates the exhaustive point correspon-
dences into the variational model. The supplementary matchings will help
the optical flow estimation to avoid most of the local minima of the original
problem and will produce a better local minima capturing large displace-
ments. The proposed confidence measure is used to determine the spe-
cific location where the movement estimation could be improved using the
matchings computed by exhaustive search. Confidence measures have been
previosly used in the literature to validate the correctness of a given optical
flow field at each point and we refer to Chapter 12 for a review.

Specific location for exhaustive search

Let us assume that, from a given sequence of frames, we have already com-
puted an occlusion mask χ and an optical flow. We assume that they have
been computed using by minimizing the energy (13.24) proposed in previ-
ous section. This is why we will denote the optical flow by u_{rc} (standing
for optical flow for color images and robust to brightness changes). Let us
denote by α the obtained adaptive map. Now, in order to determine spe-
cific locations where supplementary matchings could improve the movement
estimation, we evaluate the data cost in (13.24) (equivalently, in (13.22)) at
each $x \in \Omega$ with the computed $u_{rc}(x)$ and $\chi(x)$, namely,

$$E_{d\alpha}^c(u_{rc}, \chi)(x) =$$

$$\lambda(1 - \chi(x)) \sum_{i=1}^{3} \alpha^k(x) \left| \rho_1^k(u_{rc}(x)) \right| + \lambda\chi(x) \sum_{i=1}^{3} \alpha^k(x) \left| \rho_{-1}^k(u_{rc}(x)) \right|$$

$$+ \lambda(1 - \chi(x)) \sum_{i=4}^{5} \alpha^k(x) \left| \rho_1^k(u_{rc}(x)) \right| + \lambda\chi(x) \sum_{k=4}^{5} \alpha^k(x) \left| \rho_{-1}^k(u_{rc}(x)) \right|.$$

$$(13.40)$$

The intuitive idea is that if the value $E_{d\alpha}^c(u_{rc}, \chi)(x)$ is big, then the estimation might be improved. Additionally, we consider the smaller eigenvalue $\lambda(x)$ of the structure tensor associated to the image I_0. This structure tensor is computed in a $N \times N$ block centered at x. From these values $E_{d\alpha}^c(u_{rc}, \chi)(x)$ and $\lambda(x)$ we define a set Ω_{χ_p}, made of the points x where supplementary matchings could improve the motion estimation, by

$$\Omega_{\chi_p} = \left\{ x \in \Omega \mid E_{d\alpha}^c(u_{rc}, \chi)(x) > \theta_E \wedge \lambda(x) > \theta_\lambda \right\} \tag{13.41}$$

where θ_E and θ_λ are given constants (which we will determine empirically and fix for the experiments in Chapter 14. That is, if $E_{d\alpha}^c > \theta_E$, then we assume that the error in (13.40) is large enough to be improved using a supplementary match. In this case, we ask for an enough structure, namely, if $\lambda(x) > \theta_\lambda$, we assume that the image has sufficiently structure to match two points in a reliable form.

The set of points that belong to Ω_{χ_p} define a binary mask, which we denote by $\chi_p : \Omega \to [0, 1]$ and that will be used below. The intuitive idea is that if $x \notin \Omega_{\chi_p}$ (i.e., $\chi_p(x) = 0$), we assume that the optical flow of the point is accurate enough computed using (13.24). Otherwise, if $x \in \Omega_{\chi_p}$ (i.e., $\chi_p(x) = 1$), the computation will be helped integrating it with additional matches.

A function χ_p was presented in Stoll et al. (2012) but tey do not consider an occlusion mask or a balance term $\alpha(x)$

Additional information by exhaustive search

Exhaustive search is performed to find correspondences between points of the frame I_0 (usually called reference image) and the points of the second

frame I_1 (usually called target image). To perform the exhaustive search of a point (x_1, x_2) in the reference image we consider a squared block centered at this point and of semi-lenght p (that is, a block of size $(2p+1) \times (2p+1)$ pixels) and we search for the most similar block in the target image. Figure 13.3 schematically illustrates it.
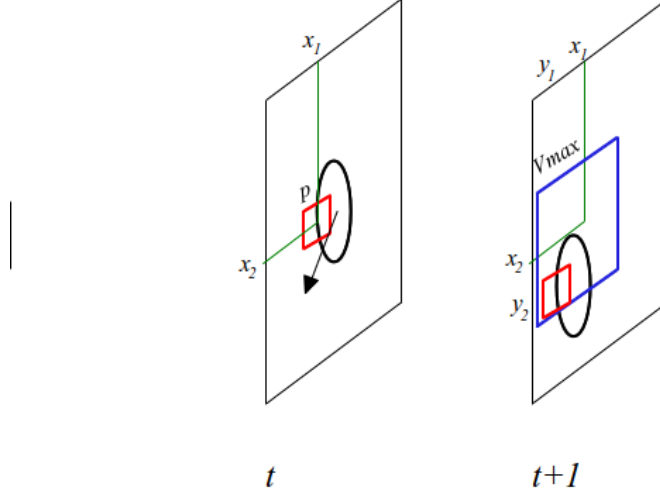
Figure 13.3: Exhaustive search schematic representation for a secuence of two frames where the white balloon moves in frame t to a new position (indicated by the black arrow) in frame $t + 1$.

Figure 13.3 shows two frames of a sequence where a white balloon moves. The movement of the balloon is represented with a black arrow in the reference image at time t. A point in the boundary of the balloon is choosed, having coordinates (x_1, x_2). A squared block of radious p around this point is indicated with red rectangle. Let denote it by B_p. Then, the search region in the target image at time $t + 1$ is illustrated in blue and defined by a squared region centered at (x_1, x_2) and with semi-lenght $v_{\max}$. That is, the size of the searching region is $(2v_{\max} + 1) \times (2v_{\max} + 1)$ pixels. Let us remark that $v_{\max}$ represents an upper bound of the expected displacement. In other words, in our algorithm $v_{\max}$ imposes a maximum displacement that is possible to recover at each point.

It is performed the sum of the absolute differences between the block B_p of the reference image at time t and each block at time $t + 1$, centered at each point in the searching area in the target image, while keeping the point (y_1, y_2) giving the minimum of these sums of the absolute differences.
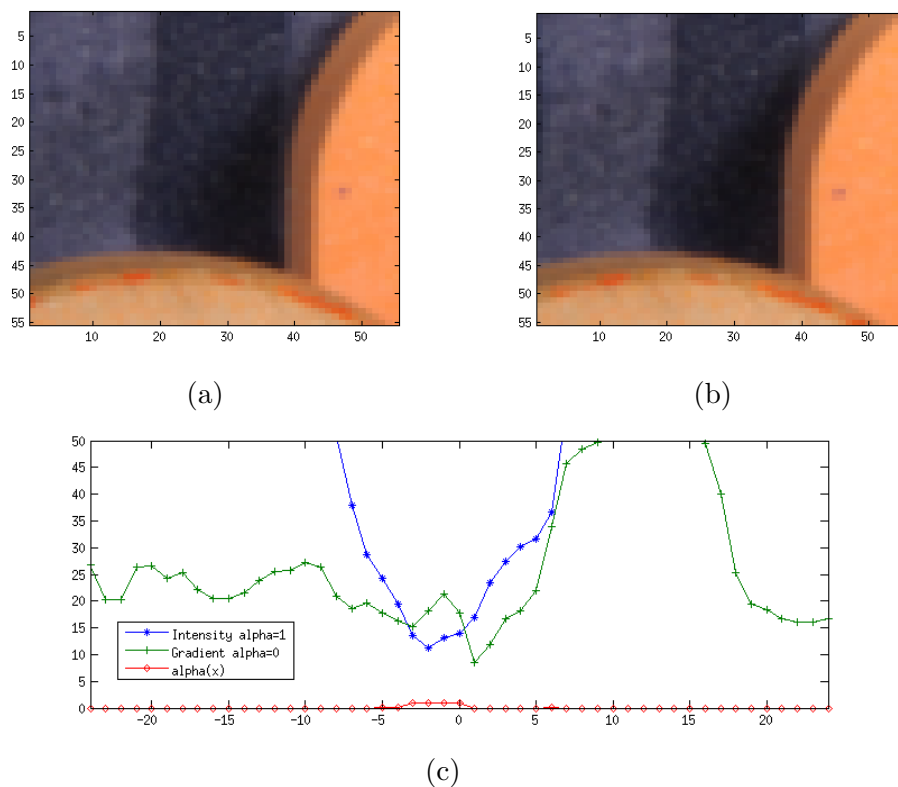
Figure 13.4: Cropped area RubberWhale sequence and Data term for intensity blocks and gradient blocks. (a) Cropped frame 10. (b) Cropped frame 11. (c) D_I and $D_{\nabla I}$ and balance term $\alpha(x)$.

In our case we propose, at each point $(x_1, x_2) \in \Omega_{\chi_p}$, to perform block matching minimizing an energy that considers the occlusion mask χ and the weight map $\bar{\alpha}(x)$. As a proof of concept we present an experiment to show the use of gradient term ($D_{\nabla I}$) and intensity(D_I) term to improve the block matching. In Figure 13.4 we show two cropped consecutives frames of Rubberwhale sequence.

Figure 13.4 shows a cropped area of frame 10 and frame 11 of RubberWhale video sequence. In Figure 13.4 (a) we took a neighborhood around the central point, we called this neighborhood B_{p1}. B_{p1} is a block of size 7x7 pixels centered in the cropped image in Figure 13.4 (a), i.e in coordinated (28,28). We search for B_{p1} in the cropped frame 11 (b) only in the horizontal direction and only in row 28. The real movement of the point (28,28) in

frame 11 is horizontal and approximately 1 pixel to the right. The presence
of shadows make the brightness constancy assumption to fail.

We have computed the sum of absolute differences of each color component
(D_I) between B_{p1} and each block in row 28 of frame 11. Additionally we
have computed the sum of absolute differences of each gradient component
($D_{\nabla I}$) between B_{p1} and each block in row 28. We present in Figure 13.4 (c)
D_I in blue, $D_{\nabla I}$ in green. We observe that the minimum of the blue curve
is located in $x = -2$ and in the other hand the minimum of the green curve
($D_{\nabla I}$) in located in $x = 1$ in the positive side of the "x" axis. Selecting the
minimum value between the minimum values of each curve, (the minimum
value of $D_{\nabla I}$) we compute correctly the displacement. Additionally we show
the $\alpha(x)$ value computed for this example. This $\alpha(x)$ value is 0 for $x = 1$
means that gradient should be used to compute the displacement.

We present a second experiment to show the use of gradient term ($D_{\nabla I}$)
and intensity (D_I) term to improve the block matching. In Figure 13.5 we
show another crop of Rubberwhale sequence.

Figure 13.5 shows the cropped area of frame 10 and frame 11 of Rubber-
Whale. In Figure 13.5 (a) we took also neighborhood around the central
point, we called this neighborhood B_{p2}. B_{p2} also is a block of size 7x7 pixels
centered in the cropped image in Figure 13.5 (a), i.e in coordinated (28,28).
We search for B_{p2} in the cropped frame 11 (b) only in the horizontal direc-
tion and only in row 28. The real movement of the point (28,28) in frame 11
is horizontal and approximately 4 pixel to the left. There is no shadows or
reflexions in this case, it means that the brightness constancy assumption
holds.

We have computed the sum of absolute differences of each color component
(D_I) between B_{p2} and each block in row 28 of cropped frame 11. Addi-
tionally we have computed the sum of absolute differences of each gradient
component ($D_{\nabla I}$) between B_{p2} and each block in row 28. We present in
Figure 13.5 (c) D_I in blue, $D_{\nabla I}$ in green. We observe that the minimum of
blue curve (D_I) is located in $x = -4$ and the minimum of the green curve
($D_{\nabla I}$) in located in $x = -4$ in the negative side of the "x" axis. Using the
minimum value between these two curves (the minimum of D_I) we compute
correctly the displacement. Additionally we show the $\alpha(x)$ value computed
for this example. This $\alpha(x)$ value is 1 for $x = -4$ means that intensity
should be used to compute the displacement.

We propose to minimize the following energy using block matching:

(a)   (b)

(c)
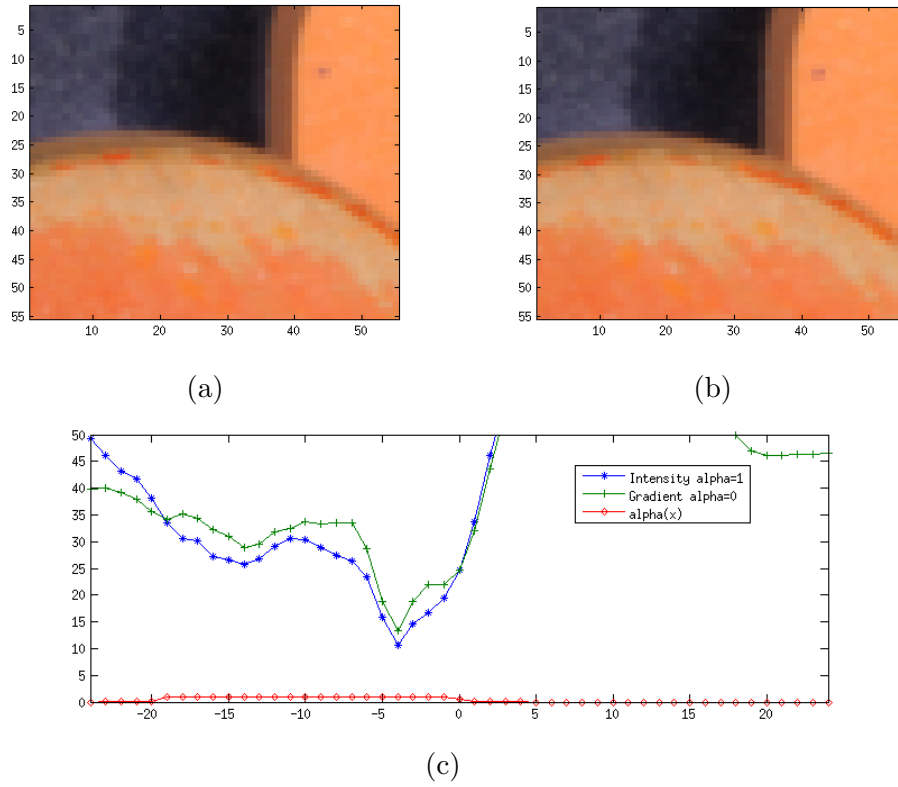
Figure 13.5: Second cropped area of RubberWhale sequence and Data term for intensity blocks and gradient blocks. (a) Cropped frame 10. (b) Cropped frame 11. (c) D_I and $D_{\nabla I}$ and balance term $\alpha(x)$.

$$E_{exha}(u_e, \chi)(x) =$$

$$\lambda(1 - \chi(x)) \sum_{k=1}^{3} \alpha^k(x) \left| I_0^k(x) - I_1^k(x + u_e(x)) \right|$$

$$+ \lambda\chi(x) \sum_{k=1}^{3} \alpha^k(x) \left| I_0^k(x) - I_{-1}^k(x - u_e(x)) \right| dx$$

$$+ \lambda(1 - \chi(x))\tau \sum_{k=4}^{5} \alpha^k(x) \left| I_0^k(x) - I_1^k(x + u_e(x)) \right|$$

$$+ \lambda\chi(x)\tau \sum_{k=4}^{5} \bar{\alpha}^k(x) \left| I_0^k(x) - I_{-1}^k(x - u_e(x)) \right| dx, \quad (13.42)$$

Let us remark that the exhaustive search is only performed in Ω_{χ_p} domain. If $E_{d\alpha}^c(u_{rc}, x) > E_{exha}(u_e, \chi(x))(x)$, we assume that previous exhaustive search may improve the solution u_{rc} at a x. If this holds, we keep $\chi_p(x) = 1$ else we modify it to $\chi_p(x) = 0$.

The vector field that minimizes (13.42) is the optical flow u_e we called a prior vector field.

Finally, we define a confidence map as follows. For each point $x \in \chi_p$, we define the confidence value $c(x)$ by

$$c(x) = \left(\frac{d_2 - d_1}{d_1} \right)^2 \left(\frac{E_{d\alpha}^c(u_{rc}, x)}{E_{exha}(u_e, x)} \right)^2 \tag{13.43}$$

where d_1, d_2 is the distance to the first and second best candidate respectively of the exhaustive search, $E_{d\alpha}^c(u_{rc}, x)$ is the error defined in (13.40) and $E_{exha}(u_e, x)$ is the error of the exhaustive search defined in 13.42.

A Confidence function with this structure was presented in Stoll et al. (2012) but they do not consider occlusions or balance term $\alpha(x)$.

13.2.4   Integration of exhaustive matchings into the variational model

Finally and as a last step, we present our variational model to handle large displacements. It is based on adding to our optical flow model in (13.24) a term $\mu \int \chi_p c(x) |u - u_e|^2$ which incorporates the matchings obtained by exhaustive search as explained above. We will refer to this additional term as large displacement prior. Let us recall that the optical flow u_e is obtained by exhaustive search using (13.42), χ_p is defined by (13.41) and μ is a constant positive value. Then, our robust color optical flow method to handle large displacement is given by the minimization of

$$E_{\alpha l}^c(u, \tilde{v}, \chi) = E_{d\alpha}^c(\tilde{v}, \chi) + E_r(u, \chi) + \frac{1}{2\theta} \int_\Omega \sum_{i=1}^5 |u - v_k|^2 \, dx$$

$$\frac{\eta}{2} \int_\Omega \chi \sum_{i=1}^3 |v_k|^2 + \beta \int_\Omega \chi \, div(u) dx + \mu \int_\Omega c(x) \chi_p |u - u_e|^2, \tag{13.44}$$

where $E_{d\alpha}^c(\tilde{v}, \chi)$ and $E_r(u, \chi)$ are defined in (13.40) and (13.3). The large displacement prior (u_e) enforces and guide u to a better local minimum

capturing large displacements. In Chapter 14 we will present several experiments evaluating the proposal.

Minimizing the Energy

As in Section 13.1, we minimize $\widetilde{E}_{\alpha l}^c$ in (13.44) by alternating among the minimization with respect to each variable while keeping the other two fixed. In particular, the minimization of $\widetilde{E_{\alpha l}}^c$ with respect to u, v^k and χ is described in the following Propositions.

Proposition 13.10. *The minimum of $\widetilde{E}_{\alpha l}^c$ with respect to $\mathbf{u} = (u_1, u_2)$ is given by*

$$u_i = \frac{\frac{1}{5}\sum_{k=1}^{5} v_k^i + \theta\mathrm{div}(g\xi_i) + \theta\beta\frac{\partial\chi}{\partial x_i} + \mu\theta u_e\chi_p c}{1 + \mu\theta u_e\chi_p c}, \quad i = 1, 2, \qquad (13.45)$$

ξ_1 and ξ_2 are computed using the following iterative scheme

$$\xi_i^{t+1} = \frac{\xi_i^t + \frac{\tau_u}{\theta}g\nabla(\frac{1}{5}\sum_{k=1}^{5} v_k^i + \theta\mathrm{div}(g\xi_i^t) + \theta\beta\frac{\partial\chi}{\partial x_i})}{1 + \frac{\tau_u}{\theta}|g\nabla(\frac{1}{5}\sum_{k=1}^{5} v_k^i + \theta\mathrm{div}(g\xi_i^t) + \theta\beta\frac{\partial\chi}{\partial x_i})|}, \quad t = 0, 1, 2, \dots$$

$$(13.46)$$

where $\xi_i^0 = 0$ and $\tau_u \leq 1/8$.

Proposition 13.11. *Assume that $\chi : \Omega \to \{0, 1\}$. The minimum of $\tilde{E}_{\alpha l}$ with respect to $\mathbf{v_k} = (v_k^1, v_k^2)$ is*

$$\mathbf{v_k} = \begin{cases} \eta_i\mathbf{u} - \mu_i\epsilon_i\alpha(x)\nabla I_i^k(\mathbf{x}^*) & \text{if} \quad \Lambda_i^k(\mathbf{u}) > \mu_i\alpha(x)|\nabla I_i^k(\mathbf{x}^*)|^2 \\ \eta_i\mathbf{u} + \mu_i\epsilon_i\alpha(x)\nabla I_i^k(\mathbf{x}^*) & \text{if} \quad \Lambda_i^k(\mathbf{u}) < -\mu_i\alpha(x)|\nabla I_i^k(\mathbf{x}^*)|^2 \\ \mathbf{u} - \epsilon_i\rho_i^k(\mathbf{u})\frac{\nabla I_i^k(\mathbf{x}^*)}{|\nabla I_i^k(\mathbf{x}^*)|^2} & \text{if} \quad |\Lambda_i^k(\mathbf{u})| \leq \mu_i\alpha(x)|\nabla I_i^k(\mathbf{x}^*)|^2, \end{cases}$$

$$(13.47)$$

when $i = 1$ and $\epsilon_1 = 1$, $\eta_1 = 1$, $\mu_1 = \lambda\theta$, $\Lambda_1^k(\mathbf{u}) = \rho_1^k(\mathbf{u})$ when $\chi = 0$, and $i = -1$, $\epsilon_{-1} = -1$, $\eta_{-1} = \frac{1}{1+\eta\theta}$, $\mu_{-1} = \frac{\lambda\theta}{1+\eta\theta}$, $\Lambda_{-1}(\mathbf{u}) = \rho_{-1}^k(\mathbf{u}) + \frac{\eta\theta}{1+\eta\theta}\mathbf{u} \cdot \nabla I_{-1}^k(\mathbf{x} + \epsilon_i\mathbf{u}_0)$ when $\chi = 1$. Additionally we create $\mathbf{x}^* = \mathbf{x} + \epsilon_i\mathbf{u}_0$.

The arguments $\mathbf{x}$ in $\mathbf{u}, \mathbf{u}_0, \mathbf{u}_e, \mathbf{c}$ are omitted.

Once all v_k are computed, we define

$$F = \lambda\left(-\alpha(x)\sum_{k=1}^{3}\left|\rho_1^k\right| - (1-\alpha(x))\sum_{k=4}^{5}|\rho_1(v_k)|\right) + \frac{1}{5}\left(\sum_{k=1}^{5}(v_k - u)^2\right),$$

$$(13.48)$$

and

$$G = \lambda \left(-\alpha(x) \sum_{k=1}^{3} \left| \rho_{-1}^{k} \right| - (1 - \alpha(x)) \sum_{k=4}^{5} |\rho_{-1}(v_k)| \right) + \frac{\eta}{5} \left(\sum_{k=1}^{5} (v_k)^2 \right).$$

$$(13.49)$$

Proposition 13.12. *Let $0 < \tau_\psi \tau_\chi < 1/8$. Given $\mathbf{u}, \mathbf{v}$, the minimum $\bar{\chi}$ of $\tilde{E}_\theta$ with respect to χ can be obtained by the following primal-dual algorithm*

$$\begin{aligned} \psi^{n+1} &= P_B(\psi^n + \tau_\psi \, g \, \nabla \chi^n) \\ \chi^{n+1} &= P_{[0,1]} \left(\chi^n + \tau_\chi \left(\operatorname{div}(g\psi^{n+1}) - \beta \, \operatorname{div}\mathbf{u} - F - G \right) \right), \end{aligned} \qquad (13.50)$$

where $P_B(\psi)$ denotes the projection of ψ on the unit ball of $\mathbb{R}^2$ and $P_{[0,1]}(r) = \max(\min(r, 1), 0)$, $r \in \mathbb{R}$.

13.2.5   Discussion on time consistency and forward-backward consistency of the proposed model

The proposed model jointly computes the optical flow and occlusions between two frames. The optical flow estimates the movement or trajectory of points from its position in the given frame to its position in the following frame. The occluded pixels, that is, the points that are visible in a given frame but not in its following frame, are modeled by a binary occlusion function so that $\chi(x) = 1$ at those points. Therefore, even if our model uses three frames, the optical flow can be interpreted as a forward optical flow and the occlusion mask can be interpreted as a forward occlusion mask.

In this section we are interested in also computing the disoccluded regions which, up to some point that we will explain here, can be interpreted as backward occluded points. Also, we are interested in introducing a certain forward-backward robustness and time consistency as well as in introducing and a certain symmetry in our model. Occluded and disoccluded points will be relate with forward and backward occlusion masks. To do so, in this section we consider more than three frames and we propose to iterate forward and backward with respect to the frames.

Let us consider a video sequence of frames $\{I(t, x)\}$, where $t \in [0, T] \subset \mathbb{R}$ stands for the time and $x \in \Omega \subset \mathbb{R}^2$ stands for the spatial variables. To simplify the presentation, we assume that $\{I(t, x)\}$ is a gray level video. We can rely on the method proposed in Section 13.1 for computing the forward optical flow, denoted here by $u^F(t, \cdot)$, from frame $I(t, \cdot)$ at time t to frame
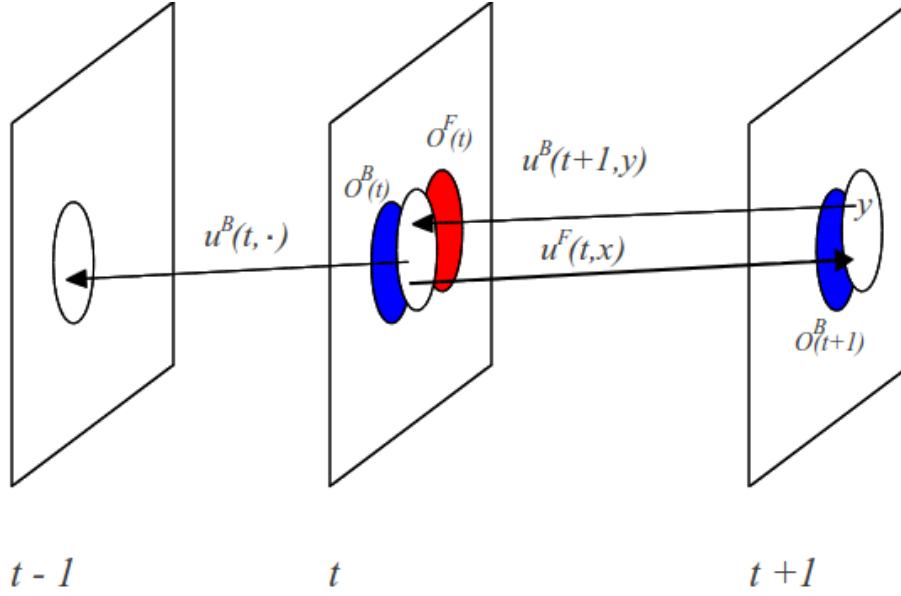
Figure 13.6: Schematic representation of three image frames used to compute the forwards and backwards optical flow and occlusion masks. In this sequence a white balloon moves from left to the right. Regions in red at t, which correspond to $O^F(t)$, are occluded at $t + 1$. Regions in blue at t, which correspond to $O^B(t)$, are disoccluded at time t but occluded at time $t - 1$ (and also regions in blue at $t + 1$ are disoccluded at $t + 1$ but occluded at time t).

$I(t + 1, \cdot)$ at time $t + 1$. The method also provides with an occlusion mask, that is, a mask of points that are visible at time t and not at time $t+1$. Let us call $\chi^F(t, \cdot)$ that forward occlusion mask. Strictly speaking $u(t, \cdot)$ is defined in $\Omega \setminus O^F(t)$, where $O^F(t)$ denotes the set $O^F(t) = \{x \in \Omega, \ \chi^F(t, x) = 1\}$. In any case, with our model we have given a sense to the flow $u(t, \cdot)$ even in $O^F(t)$. The idea is to try to better explain this sense and use it to try to also compute the disoccluded points.

Let us detail the underlying idea. First of all, let us remark that we can also use the method of Section 13.1 to also compute a backward optical flow and a backward occlusion mask, denoted here by $u^B(t, \cdot)$ and $\chi^B(t, \cdot)$, respectively. Figure 13.6 illustrates our notation on a schematic example. Regions in red (corresponding to $O^F(t)$) represent points visibles at time t but occluded at time $t + 1$. Regions in blue at time t (corresponding to $O^B(t)$) are disoccluded at time t but occluded at time $t - 1$. Analogously,

regions in blue at $t+1$ (corresponding to $O^B(t+1)$) are disoccluded at $t+1$ but occluded at time t. The vector field $u^F(t, \cdot)$ represents the forward flow from frame t to frame $t+1$ and $u^B(t, \cdot)$ the backward flow from frame t to frame $t-1$.

Actually, we would like to have $u^B(t+1, x+u^F(t,x)) = -u^F(t,x)$ for x and $y = x + u^F(t,x)$ belonging to a neither occluded nor disoccluded region. Let $\phi(t,x) = x + u^F(t,x)$ the forward map. Thus, $\phi(t, \Omega)$ represents the forward displacement of image t. The set $DO(t+1) := \Omega \setminus \phi(t, \Omega)$ are all the points that are dis-occluded at time $t+1$, that is, the points that are visible at time $t+1$ but were not visible at time t. (The $t+1$ in $DO(t+1)$ indicates that we think them as points at time $t+1$). Now, let us define $w(t+1, y) = -u^F(t, \phi(t,x))$ for $y = \phi(t,x)$ (i.e. we have reversed the flow $u^($t,x$)$) and we have a flow going from frame $t+1$ to frame t). Let $\phi_w(t,y) = y + w(t,y)$. We could define $O(t) := \Omega \setminus \phi_w(\Omega)$ (the t in $O(t)$ indicates that we think them as points at time t). This should correspond to points which are visible at frame t (background at frame t) and not visible at frame $t+1$. They should be the same as $O^F(t) = \{x \in \Omega, \ \chi^F(t,x) = 1\}$. Also, we would like to have $DO(t+1) = O^B(t+1) := \{x \in \Omega, \ \chi^B(t+1,x) = 1\}$ and $w(t+1, y) = u^B(t+1, y)$ for $y = \phi(t,x)$. Therefore, we propose an iterative method that enforces it.

Proposed method

In order to enforce to expected equalities

$$u^F(t,x) = -u^B(t+1, x + u^F(t,x)) \qquad \text{for } x \in \Omega \setminus O^F(t)$$
$$u^F(t,x) = -u^B(t,x) \qquad\qquad \text{for } x \in O^F(t) \qquad (13.51)$$

and also

$$u^B(t,x) = -u^F(t-1, x + u^B(t,x)) \qquad \text{for } x \in \Omega \setminus O^B(t)$$
$$u^B(t,x) = -u^F(t,x) \qquad\qquad \text{for } x \in O^B(t), \qquad (13.52)$$

we propose to iterate the minimization of the following pair of energies for computing $u^F(t, \cdot), \chi^F(t, \cdot), u^B(t, \cdot)$ and $\chi^B(t, \cdot)$, for all values of $t \in [0, T]$, namely,

(forward computation)

$$E^F(u, \chi, v) = E_d(v, \chi) + E_r(u, \chi) + \frac{\eta}{2} \int_\Omega \chi |v|^2 \, dx + \beta \int_\Omega \chi \text{div}(u) \, dx$$
$$+ \frac{1}{2} C_1 \int_\Omega (1-\chi) \|u(t,x) + u^B(t+1, x+u(t,x))\|^2 + \frac{1}{2} C_2 \int_\Omega \chi \|u(t,x) + u^B(t,x)\|^2,$$

$$(13.53)$$

and define

$$(u^F(t, \cdot), \chi^F(t, \cdot), v^F(t, \cdot)) = \arg \min_{(u, \chi, v)} E^F(u, \chi, v).$$

(backward computation)

$$E^B(u, \chi, v) = E_d(v, \chi) + E_r(u, \chi) + \frac{\eta}{2} \int_\Omega \chi |v|^2 \, dx + \beta \int_\Omega \chi \text{div}(u) \, dx$$
$$+ \frac{1}{2} C_1 \int_\Omega (1-\chi) \|u(t,x) + u^F(t-1, x+u(t,x))\|^2 + \frac{1}{2} C_2 \int_\Omega \chi \|u(t,x) + u^F(t,x)\|^2,$$

$$(13.54)$$

and define

$$(u^B(t, \cdot), \chi^B(t, \cdot), v^B(t, \cdot)) = \arg \min_{(u, \chi, v)} E^B(u, \chi, v).$$

for all values of $t \in [0, T]$. Previous computation is iterated until convergence. In Figure **??**, Chapter 14, we show the forward and backward masks for some frames of a sequence of the public Middlebury dataset.

Figure 13.7 displays the forward occlusion mask $\chi^F(t, \cdot)$ and the backward occlusion mask $\chi^B(t, \cdot)$, with $t = 10$ for the Urban2 sequence. Notice that the set of points x such that $\chi^F(t, x) = 1$ represent pixels which are visible in frame 10 but not in frame 11. On the other hans, the set of points x such that $\chi^B(t, x) = 1$ represent pixels which are visible in frame 10 but not in frame 9, i.e., some of the points which are disoccluded in frame 10.

Additional remarks

We would like to remark that this proposal needs still some work. Indeed, in practice these sets $O^F(t)$ (respectively $\chi^F(t)$) and $O^B(t)$ (respectively $\chi^B(t)$) may be irregular. A possible strategy could be to regularize them with a total variation approach. For instance, let $A(t)$ be any of the sets
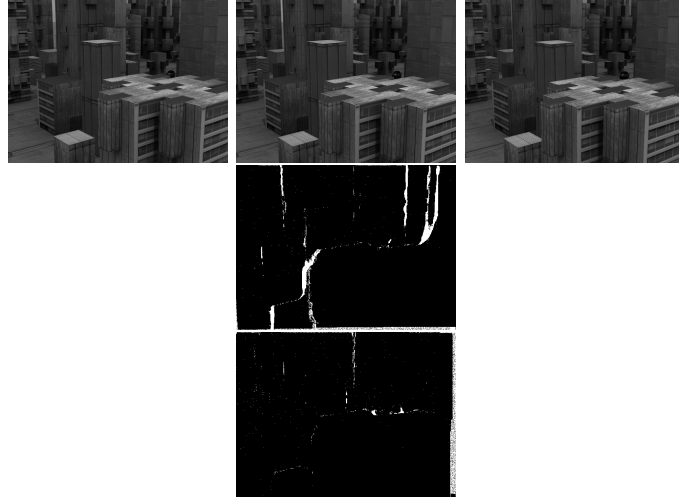
Figure 13.7: First row: From left to right, frame 9, frame 10 and frame 11 of the Urban2 sequence. Second row: forward occlusion mask $\chi^F(t, \cdot)$, with $t = 10$ containing, in white, the estimated set of pixels which are visible in frame 10 but not in frame 11. Third row: backward occlusion mask $\chi^B(t, \cdot)$ containing, in white, the set of pixels which are visible in frame 10 but not in frame 9 (that is, some of the points which are disoccluded in frame 10).

$O^F(t), O^B, DO(t + 1)$. Let $H : \Omega \to [0, 1]$ be a function representing an indicator function in a fuzzy way. A possibility could be to minimize:

$$\min_H \int_\Omega |DH| + \lambda \int_\Omega |H - \chi_{A(t)}| \qquad (13.55)$$

This represents a denoising of $A(t)$. The level sets of H give solutions.

13.3 Algorithm

This section is devoted to present the numerical algorithm for the minimization of (13.6), including pseudo-codes describing its main steps. In particular, Algorithm 13.8 present the steps of the numerical algorithm for the jointly computing the optical flow and occlusions by minimizing the model (13.6) presented in Section 13.1.

On the other hand, Algorithm 13.9 summarizes our illumination changes and large displacement robust optical flow model presented in section 13.2.3. The value of $\alpha(x)$, for all $x \in \Omega$ are updated after each propagation of the

optical flow to the finer scale, before starting the estimation of the flow field at that scale.

The data attachment $\int c(x)\chi_p(u - u_e)^2$ depends on the confidence value $c(x)$, the mask χ_p and exhaustive matchings u_e. The confidence value is an estimation of the reliability of the exhaustive matchings. The confidence value assign a higher weight to those matchings that presents strong decrease of the error.

> **Input** : Three consecutive gray level frames I_{-1}, I_0, I_1
> **Output**: Flow field $\mathbf{u}$ and occlusion layer χ for I_0
>
> Compute down-scaled images I_{-1}^s, I_0^s, I_1^s for $s = 1, \dots, N_{\text{scales}}$;
> Initialize $\mathbf{u}^{N_{\text{scales}}} = \mathbf{v}_k^{N_{\text{scales}}} = 0$, and $\chi^{N_{\text{scales}}} = 0$;
> **for** $s \leftarrow N_{\text{scales}}$ **to** 1 **do**
> > **for** $w \leftarrow 1$ **to** N_{warps} **do**
> > > Compute $I_i^s(\mathbf{x} + \epsilon_i \mathbf{u}_0(\mathbf{x}))$, $\nabla I_i^s(\mathbf{x} + \epsilon_i \mathbf{u}_0(\mathbf{x}))$, and ρ_i, $i = -1, 1$;
> > > $n \leftarrow 0$;
> > > **while** $n <$ outer_iterations **do**
> > > > Compute $\mathbf{v_k}^s$ using (13.9) (Proposition 13.2);
> > > > **for** $l \leftarrow 1$ **to** inner_iterations_u **do**
> > > > > Solve for $\xi_i^{l+1,s}$, $i \in \{1, 2\}$, using the fixed point iteration (13.8);
> > > > **end**
> > > > Compute $\mathbf{u}^s$ using (13.7) (Proposition 13.1)
> > > > **for** $m \leftarrow 1$ **to** inner_iterations_χ **do**
> > > > > Solve for χ^{m+1} using the primal-dual algorithm (13.10);
> > > > **end**
> > > **end**
> > **end**
> > If $s > 1$ **then** scale-up $\mathbf{u}^s, \mathbf{v}^s, \chi^s$ to $\mathbf{u}^{s-1}, \mathbf{v}^{s-1}, \chi^{s-1}$;
> **end**
> $u = u^1$ and $\chi = T_\mu(\chi^1)$

Figure 13.8: Algorithm for joint optical flow and occlusion computation

Input : Three consecutive color frames I_{-1}, I_0, I_1 and u_e
Output: Flow field $\mathbf{u}$ and occlusion layer χ for I_0, and $\alpha(x)$

Compute down-scaled images I_{-1}^s, I_0^s, I_1^s for $s = 1, \ldots, N_{\text{scales}}$;
Initialize $\mathbf{u}^{N_{\text{scales}}} = \mathbf{v}_k^{N_{\text{scales}}} = 0$, and $\chi^{N_{\text{scales}}} = 0$, $\alpha^{N_{\text{scales}}}(x) = 1.0$,
$\gamma = 0$;
for $s \leftarrow N_{\text{scales}}$ **to** 1 **do**

> Compute $\alpha^s(x)$ using (13.28);
> **If** $s = 1$ **then** $\gamma = 1$;
> **for** $w \leftarrow 1$ **to** N_{warps} **do**
>
>> Compute $I_i^s(\mathbf{x} + \epsilon_i \mathbf{u}_0(\mathbf{x}))$, $\nabla I_i^s(\mathbf{x} + \epsilon_i \mathbf{u}_0(\mathbf{x}))$, and ρ_i, $i = -1, 1$;
>> $n \leftarrow 0$;
>> **while** $n <$ outer_iterations **do**
>>
>>> Compute $\mathbf{v_k}^s$ using (13.47) (Proposition 13.11);
>>> **for** $l \leftarrow 1$ **to** inner_iterations_u **do**
>>>
>>>> Solve for $\xi_i^{l+1,s}$, $i \in \{1, 2\}$, using the fixed point iteration (13.46);
>>>
>>> **end**
>>> Compute $\mathbf{u}^s$ using (13.45) (Proposition 13.10) considering data attachment $\mu \int c(x) \chi_p(u - u_e)$, if $\gamma = 1$;
>>> **for** $m \leftarrow 1$ **to** inner_iterations_χ **do**
>>>
>>>> Solve for χ^{m+1} using the primal-dual algorithm (13.50);
>>>
>>> **end**
>>
>> **end**
>
> **end**
> **if** $\gamma = 1$ **then**
>
>> Compute $E_{d\alpha}^c(x)$, $\lambda(x)$;
>> Compute $\chi_p(E_{d\alpha}^c(x), \lambda(x), \theta_\lambda, \theta_E)$ implies Ω_{χ_p};
>
> **end**
> **If** $s > 1$ **then** scale-up $\mathbf{u}^s, \mathbf{v}^s, \chi^s$ to $\mathbf{u}^{s-1}, \mathbf{v}^{s-1}, \chi^{s-1}$;

end
$u = u^1$ and $\chi = T_\mu(\chi^1)$

Figure 13.9: Algorithm for illumination changes and large displacement robust optical flow

Results

In this chapter we present quantitative and qualitative results obtained from the different optical flow estimation models presented in Chapter 13. The proposed methods have been tested in two publicly available databases, namely, Middlebury database Baker et al. (2011) and MPI Sintel database Butler et al. (2012).

The chapter is organized as follows. First, we present in Section 14.1 optical flow results obtained by using the model presented in Section 13.1 for the gray level images of Middlebury database. Then, in Section 14.2 we present results obtained with the model proposed in Section 13.2.1 for the color images of Middlebury database. In section 14.3, we present results on the Middlebury database of our model considering a weight map to balance the contribution of intensity and gradient, corresponding to the energy model of Section 13.2.2. Finally, Sections 14.4, 14.5 and 14.6 present results from our complete model including a new term to handle large displacement and proposed in Section 13.2.3.

14.1 Jointly Optical Flow and Occlusion Estimation

We first evaluate the model proposed in Section 13.1, that jointly estimates occlusions and optical flow by using three consecutive frames, on the Middlebury database. Thus we consider the Middlebury sequences having at least three frames. Figure 14.1 and Figure 14.2 show some of these consecutive frames. In order to better illustrate the performance of the different energies, we have chosen to divide the Middlebury sequences in two groups.

The first one, made of all of those sequences where the displacements are smaller than the size of the corresponding moving scene objects and having no major illumination changes (see Figure 14.1). The second group is made of Middlebury sequences containing displacements larger than the size of the object itself, shadows or reflections (Figure 14.2).

Each row of each figure shows three consecutive frames, namely, frame 9, frame 10 and frame 11. In our model, I_1 corresponds to frame 9, I_0 to frame 10 and I_1 to frame 11. In the first sequence of Figure 14.1, Grove2, the camera moves in the vertical direction. The movement of the camera produces occlusion between the tree and rocks in the background. In the second sequence, Grove3, the camera moves in diagonal direction to the right bottom corner of the image. The movement of the camera occludes rocks, leaves and branches. In the Rubberwhale sequence there are movements fronto-parallel to the camera, actually to the right (e.g., curtains move to the right) and the wheel rotates and moves to the left. There are occlusions between the "z" letter and the wheel and also between curtains and the whale. In the Hydrangea sequence while camera moves to the left the flower presents a rotation on its own axis. The flower rotate from right to left. There are occlusions between the petals and leaves of the flower. In Urban2, the camera moves to the right. The apparent movement of the buildings produces occlusions between them. In sequence Urban3, camera moves in the vertical direction, and the apparent movements of the buildings produces occlusions oriented in the vertical direction. Let us notice that in all those sequences in Figure 14.1, the displacements are smaller than the size of the corresponding moving scene objects.

In Figure 14.2 we show images of the second set of Middlebury dataset. Two of these natural sequences contain displacements larger than the size of the object in the scene. The other two sequences, although presenting small displacements, contain shadows and reflections. The first row of Figure 14.2 shows three consecutive frames of the Beanbags sequence. The man juggles three balls. The camera keeps static and the balls move. Let us observe that shadows moves on the T-shirt. The second row shows DogDance sequence. In the sequence there is blur in the feet of the girls and camera movement to the left. In sequence Minicooper there are lights and shadows that move on the face of the man, and the sky reflects on the glass. The sequence Walking presents a man that project a big shadow on the wall. There is an apparent movement due to shadows on the wall.

Figures 14.4 and 14.5 show the obtained optical flow and the estimated
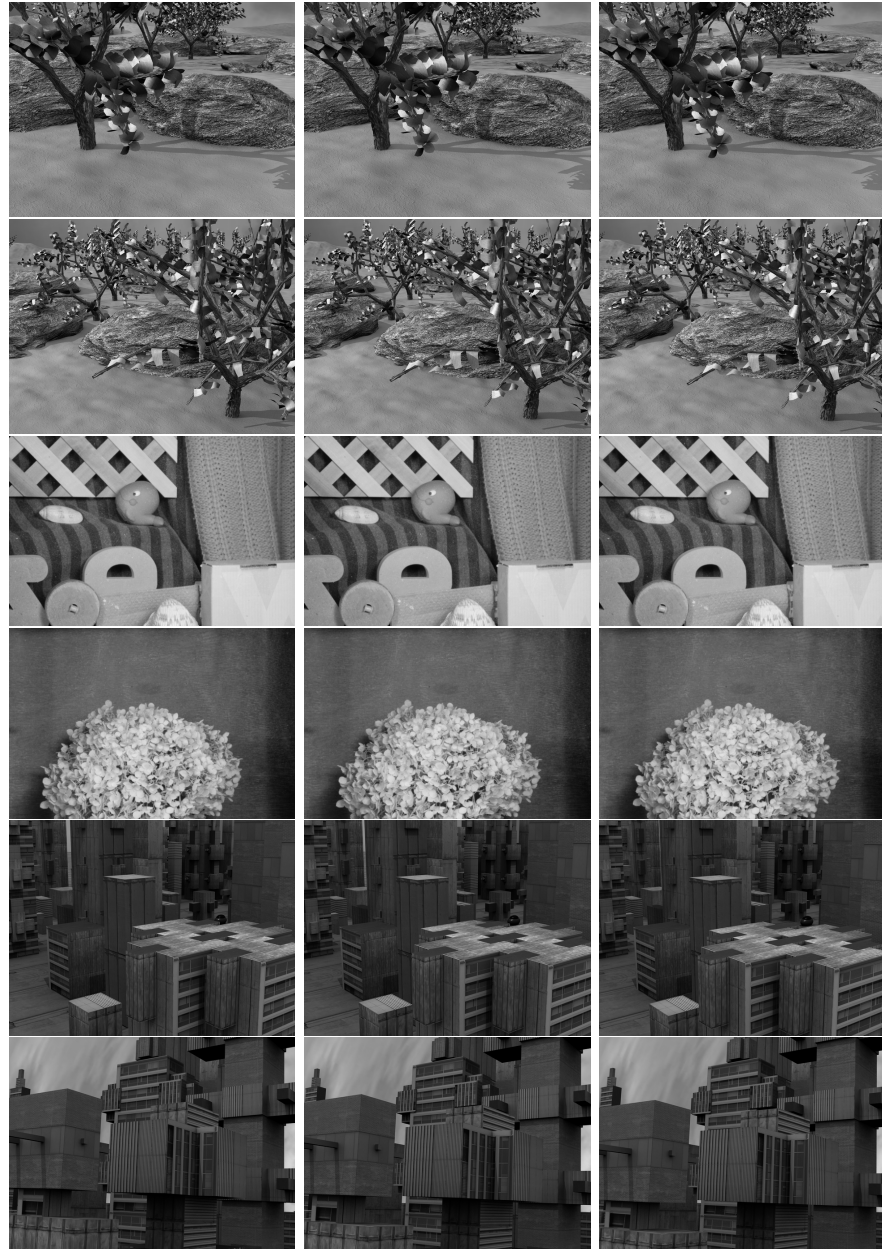
Figure 14.1: Gray level images of the Middlebury database containg small displacements. First column frame 09, second column frame 10 and third column frame 11. From top to bottom. First row , Grove2 sequence. Second row: Grove3 sequence. Third row: Rubberwhale sequence. Fourth row: Hydrangea sequence. Fifth row: Urban2 sequence. Sixth row: Urban3 sequence.

Figure 14.2: Large Displacement images of the Middlebury database, some of them containing illumination changes, shadows and light reflexions. From top to bottom: sequence Beanbags, Dogdance, Minicooper, Walking.
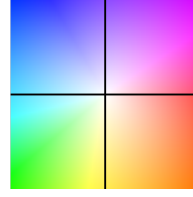
Figure 14.3: Encoding optical flow values.

occlusion mask. In order to display the optical flow results, we encode the direction and magnitude of the optical flow using the color coding scheme of Figure 14.3.

Figure 14.4 show our results for the first set of sequences with small displacements of Figure 14.1. The ground truth is shown in the first column using the same color coding of Figure 14.3. The middle column shows the estimated optical flow. The third column shows the estimated occlusion layer. Let us notice that the optical flow is correctly estimated in four of the sequences: Grove2, Grove3, Urban2 and Urban3. The presence of shadows in the real sequences Rubberwhale and Hydrangea cause that the optical flow estimation fails.

As for the occlusions, in the case of Rubberwhale and Hydrangea sequences the provided ground truth of Middlebury dataset includes the ground thruth occlusion layer, which is indicated on the optical flow ground truth image as a black color region. By comparing our estimated occlusion layer with the ground truth occlusion, we observe some differences on the wall in Hydrangea and on the "Z" in Rubberwhale. On the other hand, for the case of Urban2 and Urban3 sequences, our method is able to properly estimate the areas of image I_0 that get occluded at frame I_1 due to apparent movements of the buildings (see the corresponding occlusion masks in the last two rows of Figure 14.4). In the case of Grove2 and Grove3 the method estimate the occlusion layer with (qualitatively) high precision.

The results for the second set of Middlebury sequences are presented in Figure 14.5. The central reference image I_0 (frame 10) is shown in the first column. In the middle column we show the estimated optical flow and, in the third column, the occlusion layer estimated for each sequence. The figure shows that the optical flow for Beanbags is incorrectly estimated due to the fast movements of the balls in the sequence, which are movements larger that the size of the object itsel (the balls). In DanceDog video sequence the optical flow seems correctly estimated. In sequence Minicooper we observe
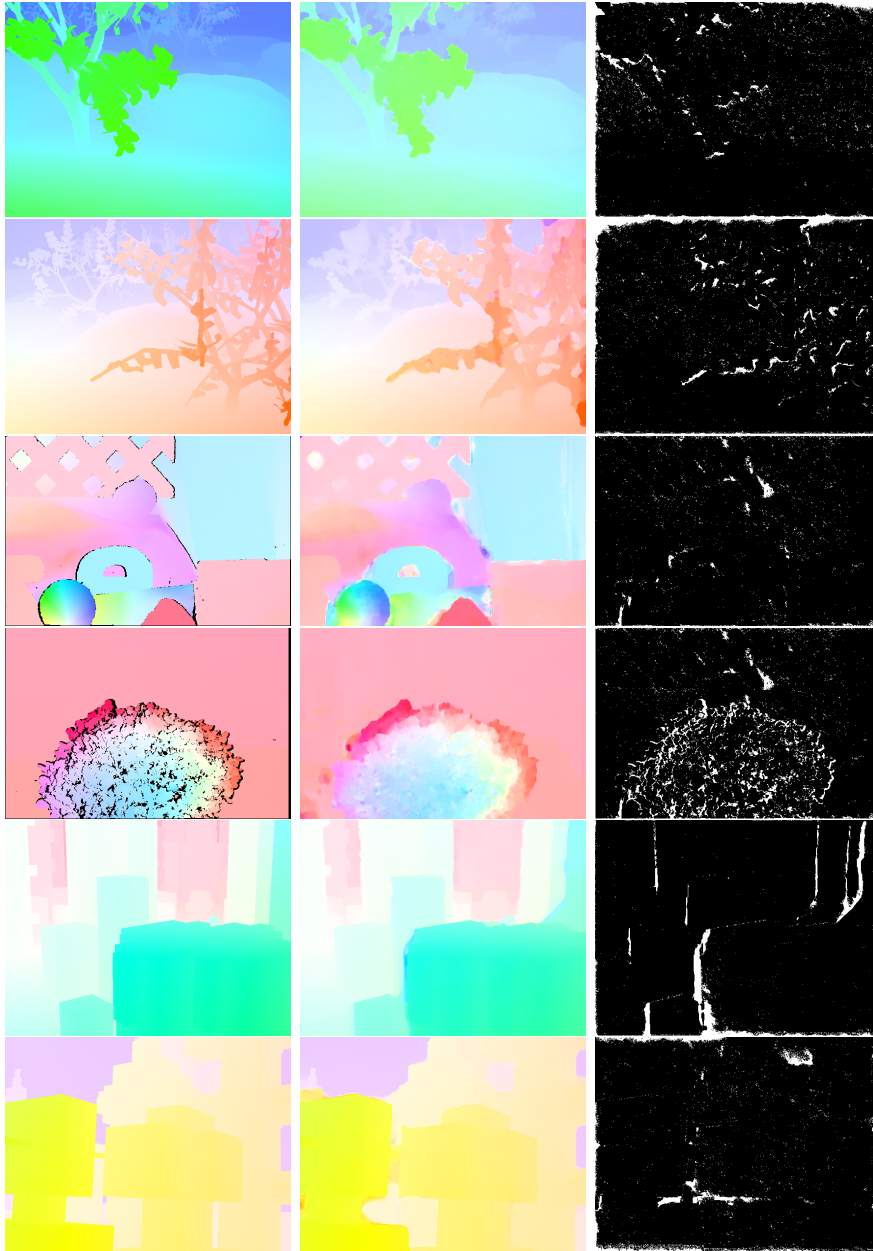
Figure 14.4: Optical flow and occlusion results obtained by our model (13.4). The results are presented using he color coding scheme of Figure 14.3. Results for Groove2, Groove3, Rubberwhale, Hydrangea, Urban2 and Urban3 sequences of Figure 14.1. First column: reference image. Second column: estimated optical flow. Third column: estimated occlusion mask.
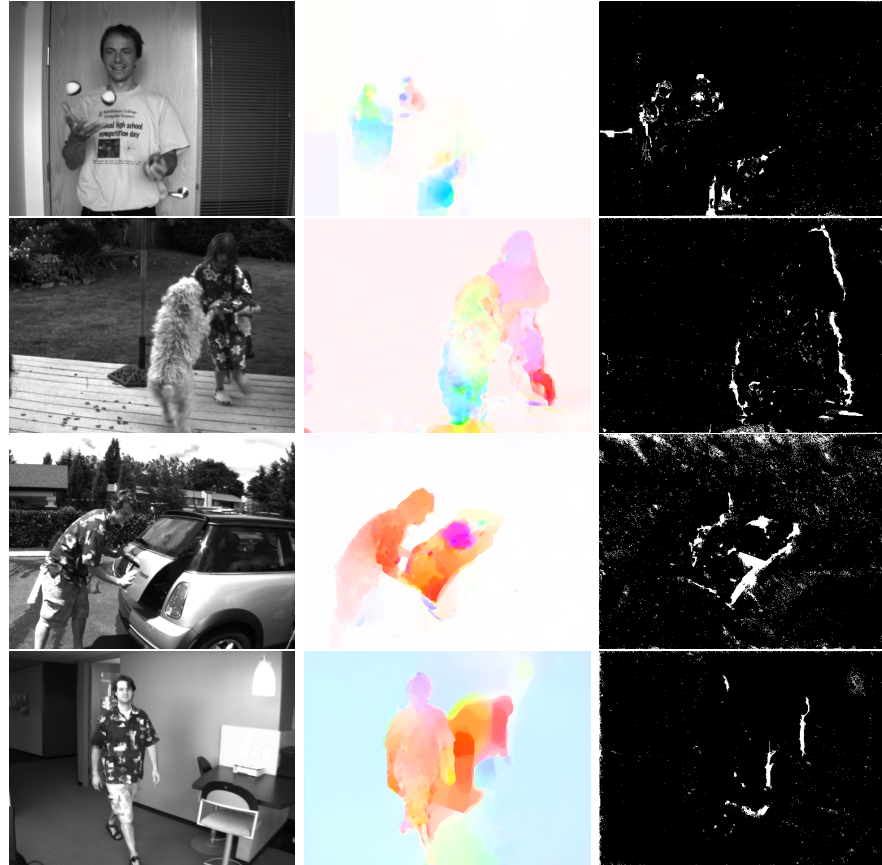
Figure 14.5: Optical flow and occlusion results obtained by our model (13.4). The results are presented using he color coding scheme of Figure 14.3. Results for Beanbags, DogDance, Minicooper and Walking sequences of Figure 14.2. First column: reference image. Second column: estimated optical flow. Third column: estimated occlusion mask.

Table 14.1: Performance of our optical flow method, on the Middlebury gray level images, based on the End Point Error (EPE) and the Average Angular Error (AAE) of the computed optical flow.

	Grove2	Grove3	Urban2	Urban3	RubberWhale	Hydra
$EPE(g)$	0.121	0.547	0.311	0.382	0.164	0.253
$AAE(g)$	1.802	5.400	2.497	3.382	5.326	2.932

a movement on the glass of the car door due to the reflections. In Walking sequence we observe a false motion detection due to the shadow on the wall.

In the case of Beanbags there are some occlusions which seems correctly estimated: on the face of the man and the occlusion produced by the right hand of the man. DanceDog sequence shows occlusions correctly estimated. The girl moves to the right in the frame I_1 and the dog moves to the left. In Minicooper there is false occlusions on the glass of the door and in other parts on the door. In the case of Walking sequence the occlusion mask presents error due to the shadow that moves.

We have also quantitatively evaluated the performance of the proposed model in this database. To this aim, all the results have been obtained with the same set of parameters, namely, $\lambda = 0.25$, $\theta = 0.30$, $\beta = 1$, $\alpha = 0$, $g = \frac{1}{1+\gamma|\nabla I_0(x)|}$, $\gamma = 0.05$, $n_{scales} = 5$, $n_{warpings} = 10$. As usual, the evaluation is based on the End Point Error and the Average Angular Error which are defined as:

$$EPE = \frac{1}{n} \sum_{i=1}^{n} \sqrt{(g_{1i} - u_{1i})^2 + (g_{2i} - u_{2i})^2}$$

$$AAE = \frac{1}{n} \sum_{i=1}^{n} cos^{-1} \left(\frac{1 + g_{1i}u_{1i} + g_{2i}u_{2i}}{\sqrt{1 + g_{1i}^2 + g_{2i}^2}\sqrt{1 + u_{1i}^2 + u_{2i}^2}} \right),$$

(14.1)

where n in the number of visible points of the image, $g = (g_1, g_2)$ is the ground truth and $u = (u_1, u_2)$ is the computed optical flow. Table 14.1 shows the End Point Error and the Angular Average Error of the estimated flow field. These two values are computed using the the occlusion layer χ, i.e. they are computed at visible pixels. Let us recall that we have evaluate our method only on those sequences having at least three frames. Optionally , we can use structure texture decomposition of the images, performing the optical flow estimation in the texture part. The use of texture images allows to improve results for real images, where it may degrade performance (with

Table 14.2: Results using structure texture decomposition: EPE and AAE optical flow results of the same Middlebury gray level images as in Table 14.1 but computed from the texture images.

	Grove2	Grove3	Urban2	Urban3	RubberWhale	Hydra
$EPE(t)$	0.121	0.547	0.311	0.382	0.093	0.162
$AAE(t)$	1.802	5.400	2.497	3.382	2.895	1.977

respect to using the original image)on synthetic images. Considering the texture component of the images we obtain in Table 14.2.

14.2 Color constancy assumption

In this section we present the results obtained with the model (13.13) that extends to color images our model (13.4) for gray level images. This model is first evaluated on the color video sequences of Middlebury database.

Figure 14.6 shows color image sequences of the Middlebury database, namely, the synthetic sequences Groove2, Groove3, Urban2 and Urban3, and the sequences RubberWhale and Hydra which contain natural scenes.

Figure 14.7 shows the obtained optical flow. The first columns shows the ground truth for each sequence. The estimated optical flow is showed in the second column, also coded with the same color scheme. The third column shows the estimated occlusion mask. Let us remark that the results in Figure 14.7 for the four synthetic sequences are similar to the corresponding ones in Figure 14.4. However, in the case of the two real sequences Rubber-Whale and Hydrangea, we obtain an improvement in the estimated optical flow. The addition of color information improve the results: in the case of Hydrangea there is less false movement due to shadows on the wall and, in the Rubberwhale case, the method is able to obtain a thinner border around the wheel than in Figure 14.4. On the other hand, the estimated occlusions seems to be more noisy although in the case of RubberWhale and Hydrangea seem to be more similar to the geometry of images. For Urban2, Urban2, Grove2, Grove3 the occlusions look thinner.

Table 14.3 shows the corresponding EPE an AAE. For comparison, we also report, in the first and second rows of the table, the values obtained by the previous gray level model (denoted with the letter (g)). The third and fourth rows in Table 14.3 correspond to our results for color images (denoted with the letter (c)). Let us notice that the quantitative results

Figure 14.6: Color sequences from Middlebury database. From left to right, frame 9, frame 10 and frame 11. These sequences present small displacements. The first row shows the Groove2 sequence, second row Groove3 sequence, third row RubberWhale, fourth row shows the Hydrangea, fifth row Urban2 and sixth Urban3 sequence.
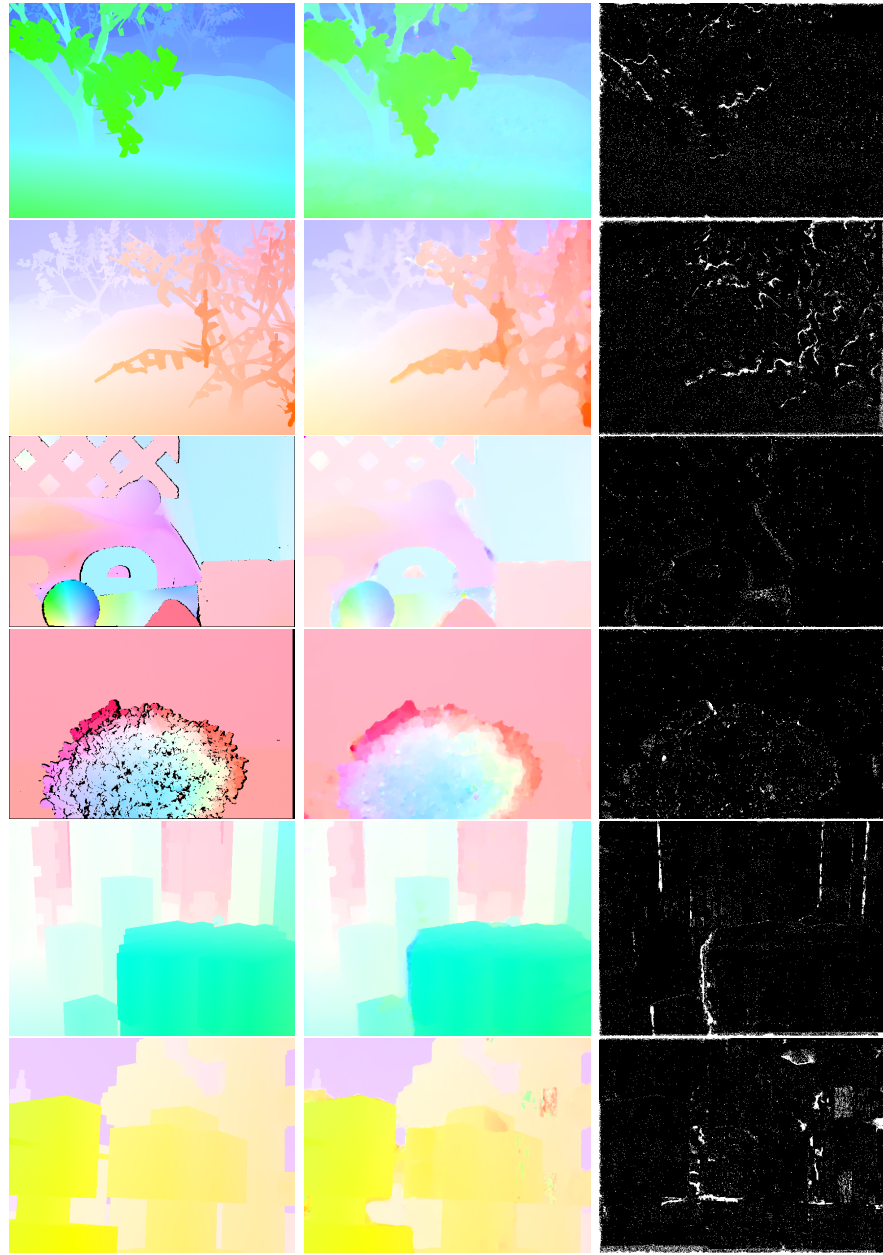
Figure 14.7: Results obtained by our extension (13.13) of the joint occlusion and optical flow model to color images.

Table 14.3: EPE and AAE of the results obtained by our extension (13.13) to color images.

	Grove2	Grove3	Urban2	Urban3	RubberWhale	Hydra
$EPE(g)$	0.121	0.547	0.311	0.382	0.164	0.253
$AAE(g)$	1.802	5.400	2.497	3.382	5.326	2.932
$EPE(c)$	0.148	0.510	0.282	0.456	0.127	0.201
$AAE(c)$	2.060	5.245	2.554	3.329	4.145	2.500

for the four synthetic are similar. In the case of real images, the results are quite different. In other word, the inclusion of colorm information improves performance (without any need of structure-texture decomposition as before).

14.3 Gradient constancy assumption

This section presents some results obtained from the model (13.33) presented in Section 13.2.2 which includes an adaptative weight map $\alpha(x)$ allowing to balance the contribution of color and gradient information.

In order to test this contribution, We performed the same experiments as in 14.2 but with different α values. First, we fix the weight map to a constant value $\alpha(x) = 0.5$ for all x. Second, we fix the weight map $\alpha(x)$ to 0.0 (which means only gradient constancy assumption) and finally we perform experiments using adaptive weight map $\alpha(x)$. The experiments in Figure 14.7 and Table 14.3 of previous section are equivalent to perform experiments of this section with a constant value $\alpha(x) = 1.0$.

14.3.1 Results obtained using a constant weight map $\alpha(x) = 0.5$

In Figure 14.8 we show the optical flow obtained with weight map $\alpha(x) = 0.5$. We present in the first column the ground truth, in the second column we present the estimated optical flow and in the last column the estimated occlusions. Some improvements appear for real images but we do not appreciate any improvements in synthetic images. Actually for Grove3 there is a wrong optical flow and occlusion estimation on the surface of two buildings.

Table 14.4 presents the numerical evaluation of these resultswith $\alpha(x) = 0.5$. By comparing this results with the ones with $\alpha = 1.0$ in Table 14.3 ,we
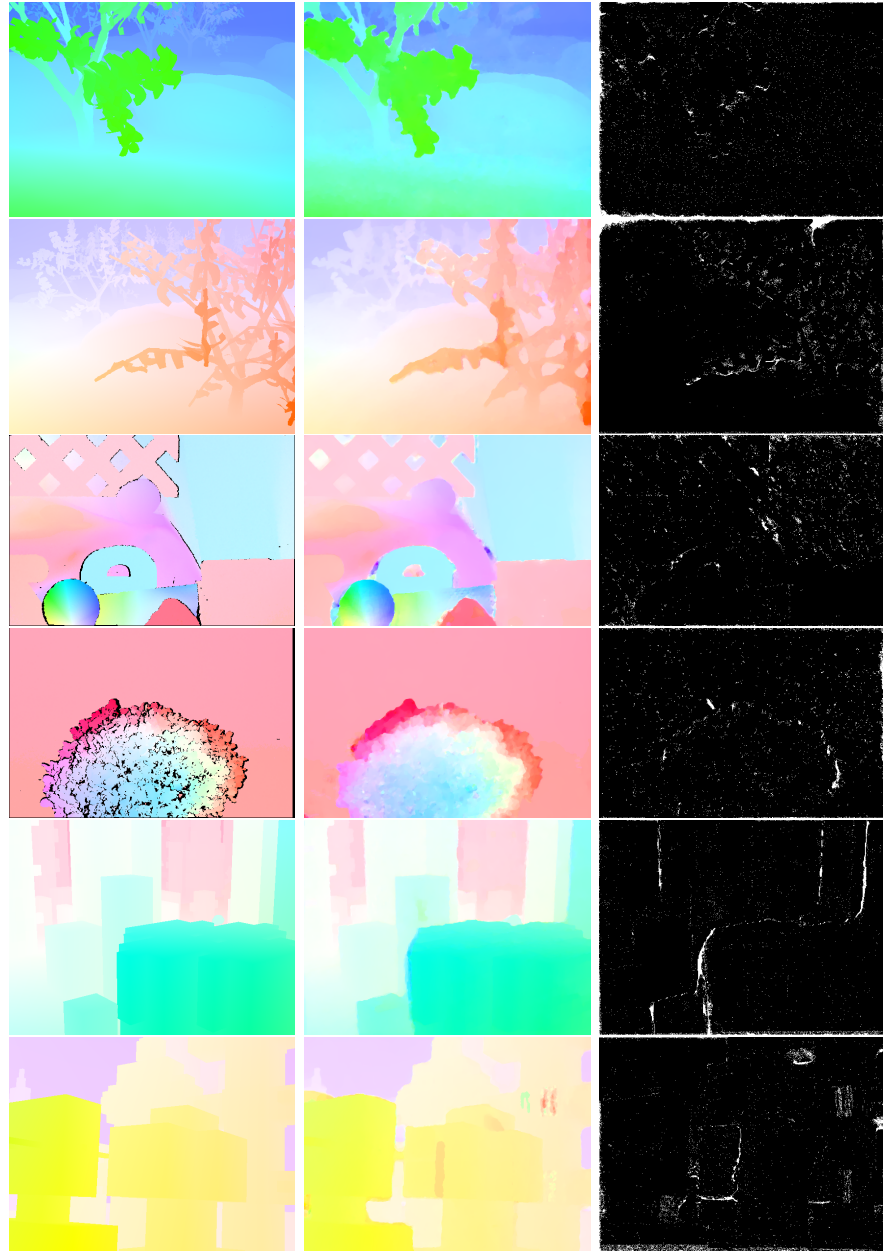
Figure 14.8: Optical flow obtained from (13.33) with $\alpha(x) = 0.5$. First column: groud truth. Second column: estimated optical flow. Third column: estimated occlusion mask. First row: Groove2 sequence. Second row: Groove3. Third row: Rubberwhale. Fourth row: Hydrange. Fifth row: Urban2 and sixth row: Urban3.

Table 14.4: EPE and AAE optical flow results from (13.33) extended to color images using $\alpha(x) = 0.5$.

	Grove2	Grove3	Urban2	Urban3	Rubber	Hydra
EPE(c) $\alpha(x) = 0.5$	0.150	0.516	0.284	0.416	0.103	0.178
AAE(c) $\alpha(x) = 0.5$	2.118	5.261	2.519	3.460	3.373	2.230

observe that now the AAE for RubberWhale drops from 4.145 to 3.373 and for Hydra it drops from 2.500 to 2.230. This tendency is presented also in the End Point Error: End point error of RubberWhale drops from 0.127 to 0.103 and for Hydrangea drops from 0.201 to 0.178. In the case of synthetic images there is no clear tendency. In Urban2, Grove2, Grove3 the AAE rises from 2.060 to 2.118, from 5.245 to 5.261 and from 3.329 to 3.460 respectively. Only in the case of Urban2 the AAE drops from 2.554 to 2.519.

This experiments reinforce the idea that the use of gradient helps to improve the estimation of the optical flow in case of real images containing reflections or shadows.

14.3.2 Results obtained using a constant weight map $\alpha(x) = 0.0$

This set of experiments is equivalent to consider only gradients in the data term. In Figure 14.9 we show the optical flow obtained by our method with constant weight map $\alpha(x) = 0.0$. Table 14.5 presents the corresponding quantitative results. We observe an reinforcement of the tendency for the values AAE an EPE. The AEE and also EPE rise for synthetics images. In the case of real images (RubberWhale and Hydrangea) the AAE has a tendency to drop. In the case of Rubberwhale AAE drops from 4.145 with $\alpha(x) = 1.0$ to 2.998 with $\alpha = 0.0$. For Hydrangea AAE drops from 0.201 to 1.969 respectively. This tendency is presented also for End Point Error. End Point Error for RubberWhale drops from 0.127 to 0.097 and for Hydrangea drops from 0.201 to 0.162. In the case of synthetic images the tendency is inverted; for instance in Grove2 the AAE rises from 0.148 to 0.173 similar for its AAE which rises from 2.060 to 2.530.
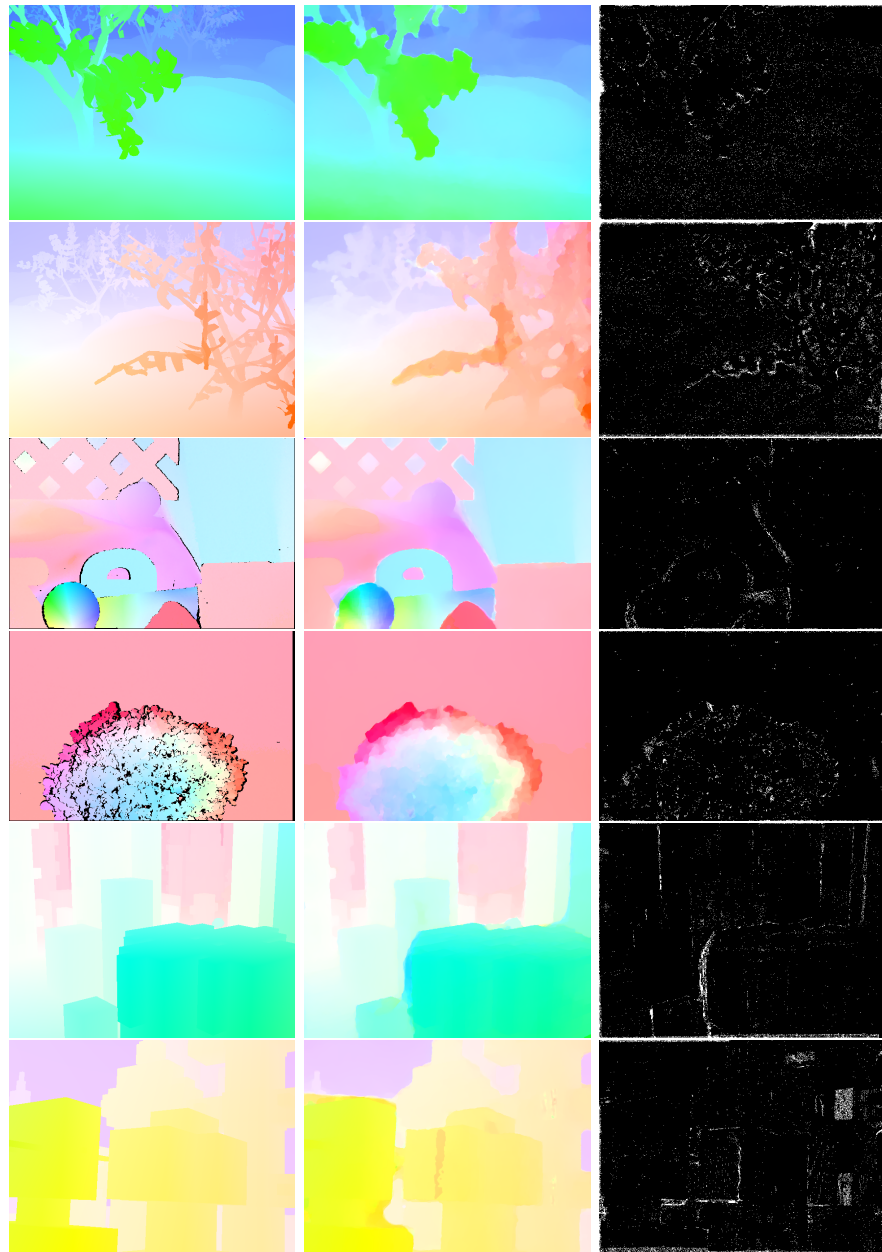
Figure 14.9: Results obtained with $\alpha(x) = 0.0$. First column: groud truth. Second column: estimated optical flow. Third column: estimated occlusion mask. First row: Groove2 sequence. Second row: Groove3. Third row: Rubberwhale. Fourth row: Hydrange. Fifth row: Urban2 and sixth row: Urban3.

Table 14.5: EPE and AAE optical flow results from (13.33) extended to color images using $\alpha(x) = 0.0$.

	Grove2	Grove3	Urban2	Urban3	Rubber	Hydra
EPE(c) $\alpha(x) = 0.0$	0.173	0.644	0.343	0.544	0.097	0.162
AAE(c) $\alpha(x) = 0.0$	2.530	6.414	2.686	4.384	2.998	1.969

Table 14.6: EPE and AAE optical flow results from (13.33) extended to color images using an adaptive $\alpha(x)$.

	Grove2	Grove3	Urban2	Urban3	Rubber	Hydra
EPE(c) $\alpha(x)$	0.163	0.630	0.301	0.486	0.092	0.158
AAE(c) $\alpha(x)$	2.350	6.306	2.525	3.787	2.830	1.943

14.3.3 Results obtained using an adaptive weight map $\alpha(x)$

In Figure 14.10 we show the optical flow obtained by our method (13.33) with an adaptive weight map $\alpha(x)$. Table 14.6, shows that using the adaptive map the AAE and EPE maintain the same tendency for the real images RubberWhale and Hydrangea. In the case of synthetic images, the use of and adaptive weight map do not decrease the AAE and EPE, comparing with the results obtained using $\alpha(x) = 1.0$. Thus we decide to kept the $\alpha(x) = 1.0$ for synthetic images and for real image use an adaptive weight map $\alpha(x)$.

We show in Figure 14.11 the color image and the adaptive weight map $\alpha(x)$ computed for RubberWhale and Hydrangea. Values of the weight map was scaled in order to be represented in a gray level scale. If the gray level of the adaptive weight map is dark means that the gradient should be used and in the other hand if the gray level is high means that intensity should be used. If we compare the color image with the weight map, we see that it is necessary to use gradient in the places with shadows and in places with edges essentially is necessary to use the intensities.

14.4 Integration of exhaustive matchings

Figure 14.12 shows the optical flow obtained by using our method (13.44) with the integration of the exhaustive matchings. We have added the additional information of the exhaustive matchings in the last scale of the multiscale pyramid. For all the experiments, the used parameters are: $\theta = 0.40$, $\lambda = 0.60$, $\alpha = 0.0$, $\beta = 1.0$, $\theta_{\lambda1} = 0.98$ and $\theta_E = 0.98$. For parameter μ we
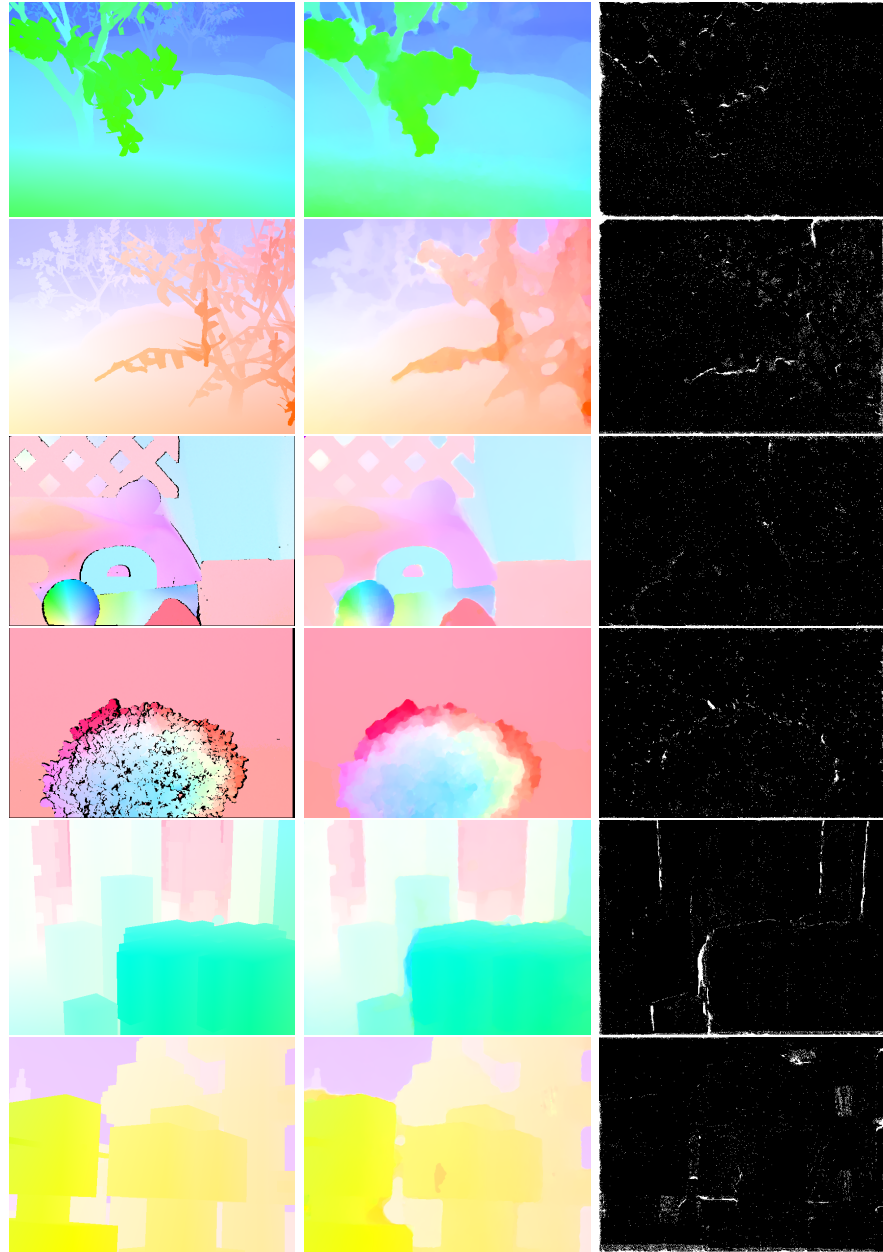
Figure 14.10: Results obtained with an adaptive $\alpha(x)$.
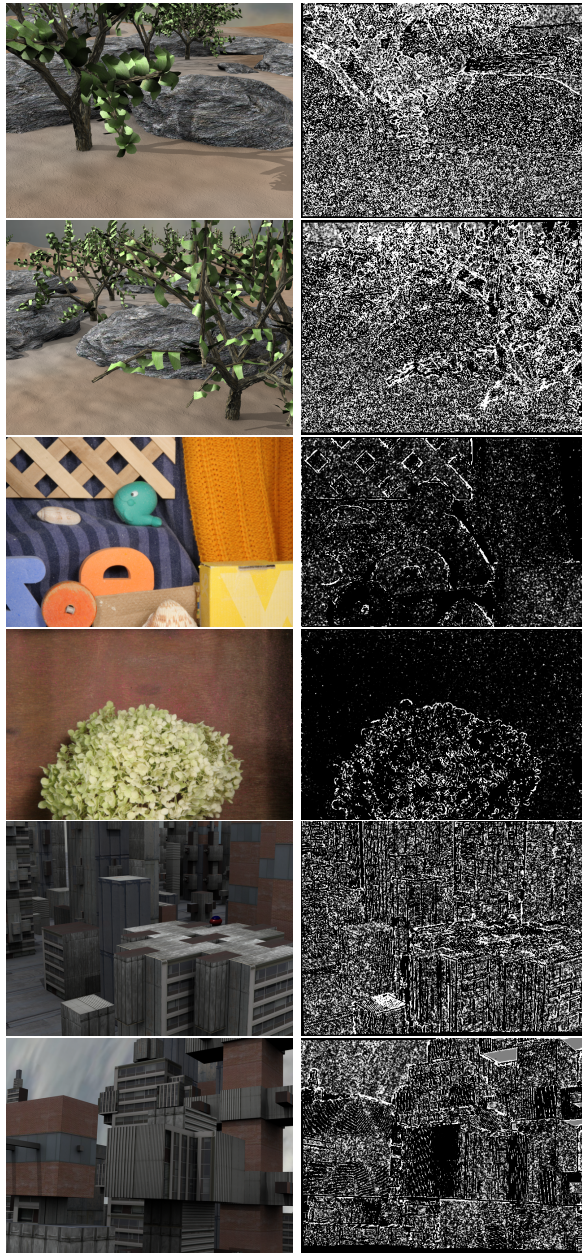
Figure 14.11: Computed adaptive weight map $\alpha(x)$ for RubberWhale and Hy-drangea. Left original image, right computed weight map.

Table 14.7: Summary of the results EPE and AAE for optical flow extended to color images using $\alpha(x) = 0.5$ + exhaustive search.

	Grove2	Grove3	Urban2	Urban3	Rubber	Hydra
$EPE(g)$	0.121	0.547	0.311	0.382	0.164	0.253
$AAE(g)$	1.802	5.400	2.497	3.382	5.326	2.932
EPE(c) $\alpha(x) = 1.0$	0.148	0.510	0.282	0.456	0.127	0.201
AAE(c) $\alpha(x) = 1.0$	2.060	5.245	2.554	3.329	4.145	2.500
EPE(c) $\alpha(x) = 0.5$	0.150	0.516	0.284	0.416	0.103	0.178
AAE(c) $\alpha(x) = 0.5$	2.118	5.261	2.519	3.460	3.373	2.230
EPE(c) $\alpha(x) = 0.0$	0.173	0.644	0.343	0.544	0.097	0.162
AAE(c) $\alpha(x) = 0.0$	2.530	6.414	2.686	4.384	2.998	1.969
EPE(c) $\alpha(x)$	0.163	0.630	0.301	0.486	0.092	0.158
AAE(c) $\alpha(x)$	2.350	6.306	2.525	3.787	2.830	1.943
EPE(c) $\alpha(x)$+ES	0.148	0.544	0.284	0.450	0.092	0.158
AAE(c) $\alpha(x)$+ES	2.083	5.539	2.560	3.430	2.801	1.937

decreased its value in each iteration. We took and initial value $\mu_o = 300$ and in each iteration update its value following the model $\mu_n = (0.6)^n \mu_0$. For real images we use blocks of 7×7 pixels and for synthetic images we used blocks of 31×31 pixels. For synthetic images we fixed $\alpha(x) = 1.0$.

Table 14.7 presents a summary of the quantitative numerical evaluations of our model for $\alpha(x) = 1.0$, $\alpha(x) = 0.5$, $\alpha(x) = 0.0$, adaptive $\alpha(x)$ and adaptive $\alpha(x)$ plus exhaustive search. We observe that for the synthetic images, the AAE values present an increment with respect to results obtained with the weight map $\alpha(x) = 1.0$ (third row), and the same happens for EPE. Only Urban3 drops from 0.456 to 0.450. Let us notice that the synthetic images present a lot of autosimilarities that causes erroneous matchings. For instance, in the Groove 3 results in Figure 14.12, there is a region where the optical flow was wrong computed due to false matchings caused by autosimilarity. On the other hand, in the case of real images we observe in the table that the performance of the method including exhaustive matchings is better.
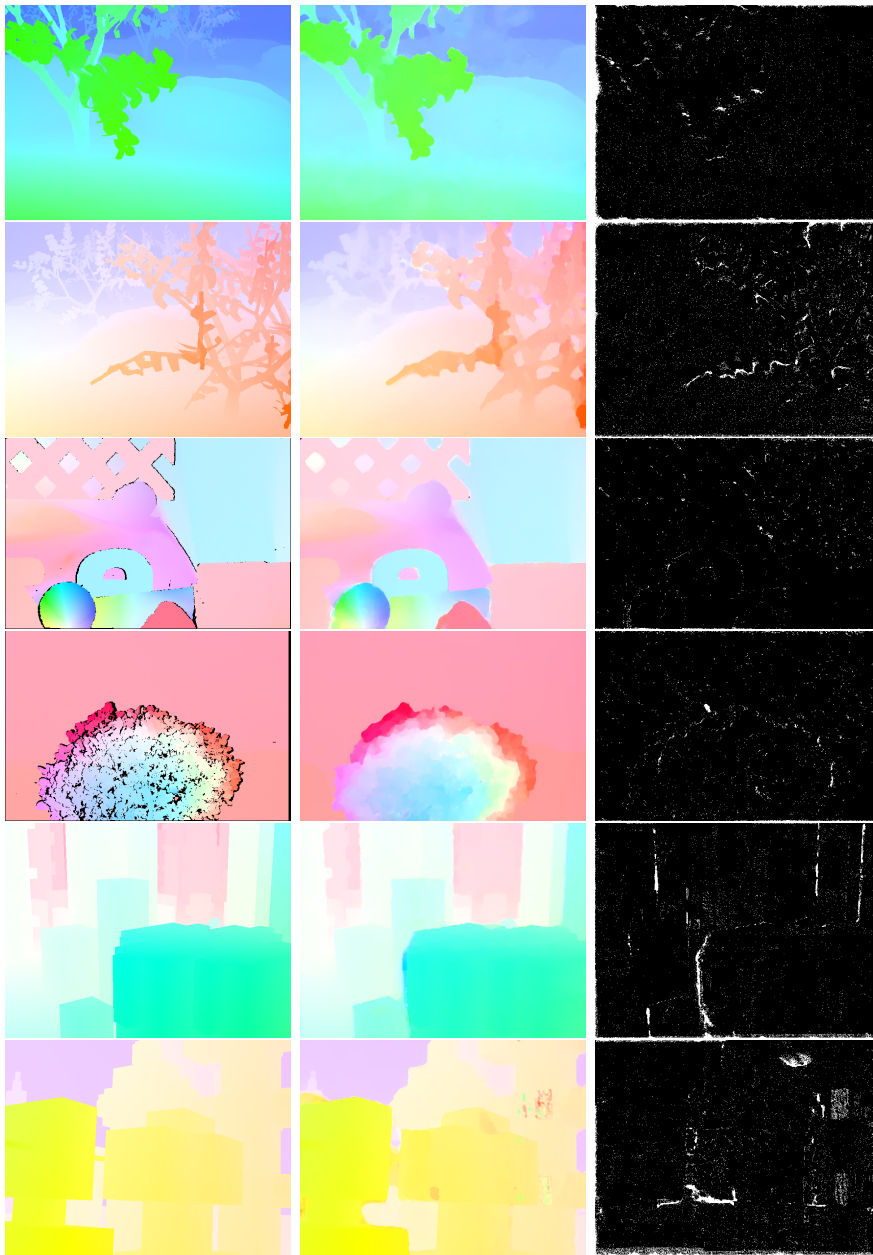
Figure 14.12: Computed optical flow integrating exhaustive matchings. First row, from left to right: optical flow for Groove2 and Groove3. Second row: Urban2 and Urban3. Third row: RubberWhale and Hydrangea

14.5 Evaluation in Large Displacement Middlebury

As commented before, the Middlebury database contains some video sequences of real scenes with large displacements, illumination changes, shadows that moves in the sequences or light reflexions. Figure 14.13 displays some of these examples, namely, the real sequences Beanbags, Basketball, Backyard and Dogdance

In these video sequences the best results where obtained when we computed the exhaustive search in every scale. We have modified the threshold θ_λ and θ_E from $\theta_\lambda = 0.98$ and $\theta_E = 0.98$ to $\theta_\lambda = 0.90$ and $\theta_E = 0.85$. The other parameter are the same as before, namely, $\lambda = 0.60$, $\theta = 0.40$, $\beta = 1$, $\alpha = 0$, $g = \frac{1}{1+\gamma|\nabla I_0(x)|}$, $\gamma = 0.05$, $n_{scales} = 5$, $n_{warpings} = 10$.

Figure 14.5 displays the obtained results for Beanbags. The maximum displacement in this video sequence is $v_{max} = 41$ pixels. We present in 14.5 (d) the estimated occlusion mask. This occlusion mask is correctly estimated thanks to the addition of exhaustive matchings that guide the optical flow. We observe that the optical flow is correctly estimated. In the center of the optical flow image (e) we observe a violet compact region that corresponds to the estimated motion of the ball of the center in (b). The violet color represents upward motion. In Figure (e) we observe a second compact region in yellow color which represents the second ball that moves downward. In (f) we show the smaller value of the structure tensor computed in I_0. The figure shows that does exist enough structure in the borders of the ball, letters on the T-shirt, on the hinge, on the doorknob, and on the door frame. The $\alpha(x)$ function (g) shows that the gradient is necessary on the hands of the man and on the T-shirt. We show in figure (h) red points indicating at which points the exhaustive flow u_e is used account to guide the optical flow estimation.

In Figure 14.15 we show the results for the DogDance sequence. We present in (d) the estimated occlusion mask. This occlusion mask is correctly estimated showing that the girl moves to the left and the dog to the right. The estimations of the occlusion is noisy but preserves correctly the structure of the image. (e) displays the estimated optical flow. By comparing the results in (e) with the ones obtained in 14.5, we can observe that the computed optical flow better captures the fast movements (e.g., the foot of the girl). In (f) we show the minimum eigenvalue of the structure tensor computed in I_0. The figure shows that does exist enough structure on the kimono

Figure 14.13: Some Middlebury real sequences containing large displacements, illumination changes, shadows that moves in the sequences or light reflexions. First column, frame 9 of the video sequences. Second column, frame 10. Third column, frame. From top to bottom: First row: BeanBags video sequences. Second row: DogDance video sequence. Third: MiniCooper video sequence. Fourth row: Walking video sequence.

of the girl, on the flowers on the floor. There is no enough texture on the shoe of the girl so the structure tensor presents a low intensity value at that points. The $\alpha(x)$ function is displayed in (g) and shows that the gradient is necessary on the girl and on the dog. We show in figure (h) the points where the exhaustive flow u_e is taken into account to guide the optical flow. For instance, there are no red points on the yellow shoe of the girl.

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)
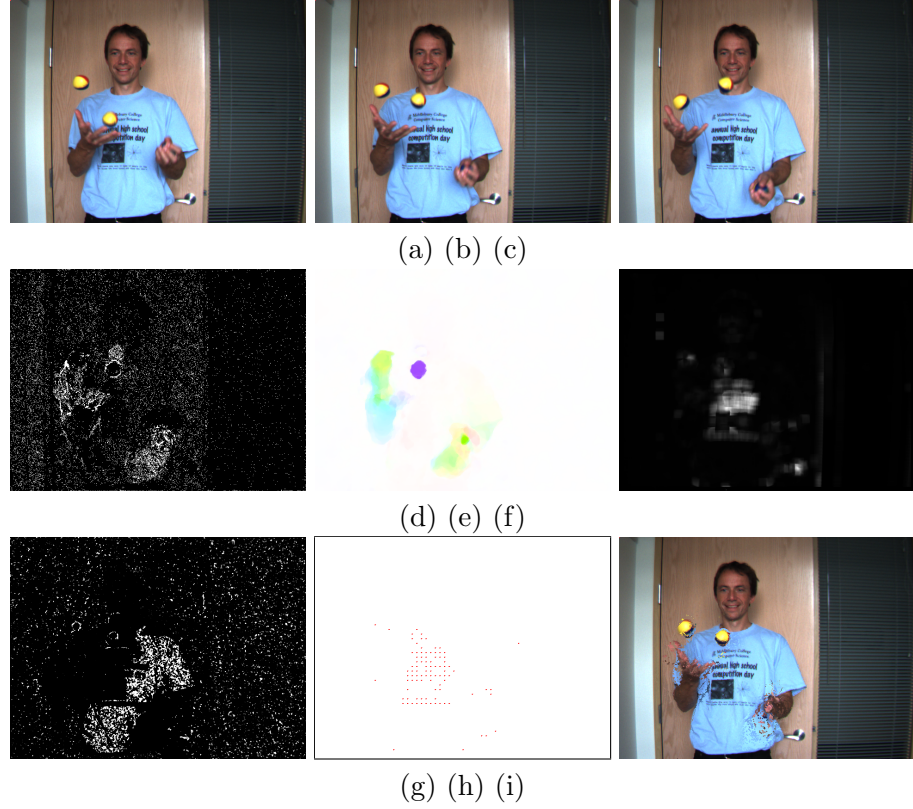
Figure 14.14: Results obtained by our proposed model in Beanbags video sequence. (a), (b) and (c) I_{-1} (frame 9), I_0 (frame 10) and I_1 (frame 11) respectively. (d) Estimated occlusion function $\chi(x)$. (e) Estimated optical flow. (f) Smaller eigenvalue of the structure tensor $\lambda(x)$. (g) Balance term between brightness gradient $\alpha(x)$. (h) Mask where X_{ip} the exhaustive flow u_e is considered to guide the optical flow estimation. (i) Compensated image.

Figure 14.16 shows the results obtained by our model for MiniCooper sequence. We present in 14.16 (d) the estimated occlusion mask showing that the door of the car moves downward and that also man moves downward. The estimations of the occlusion is noisy but preserves correctly the structure of the image. In (e) we observe the estimated optical flow. We compare our results in (e) with the one obtained in 14.5. We see that the obtained optical flow is more homogeneous (orange color) than the one presented in 14.5. There is no estimated motion on the car bumper and less estimated motion on the glass due to deflection in the sky. In (f) we show the minimum eigenvalue of the structure tensor computed in I_0. The figure shows that

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)
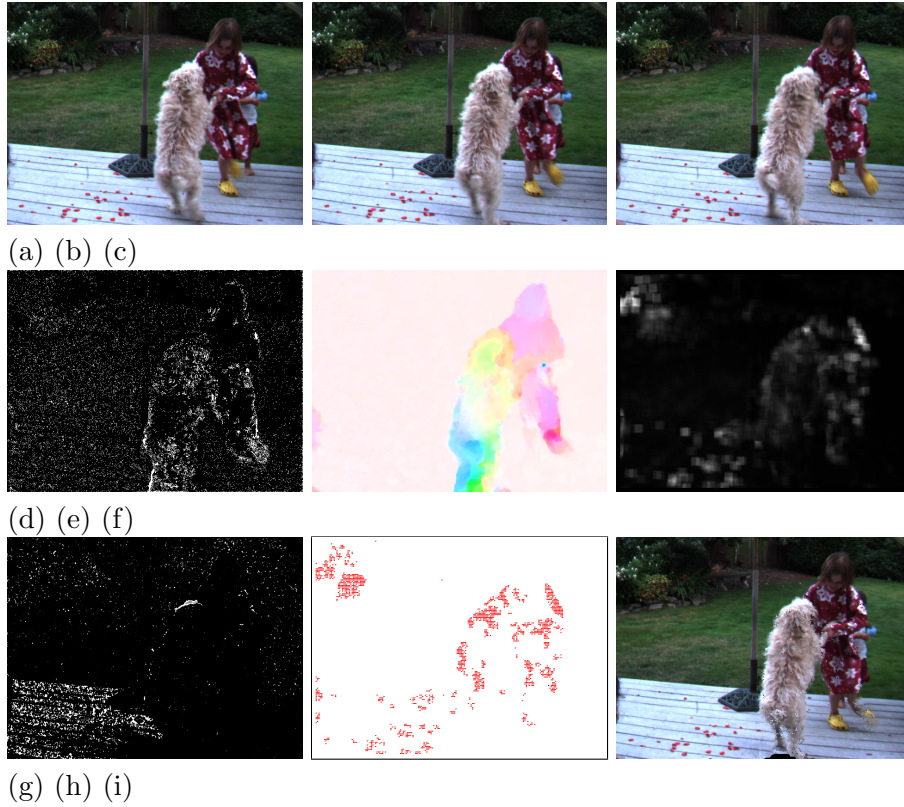
Figure 14.15: Obtained results for the DogDance video sequence. Frames 09, 10 and 11 are shown in (a), (b) and (c), respectively. (d): $\chi(x)$ function. (e): estimated optical flow. (f) $\lambda(x)$. (g) $\alpha(x)$. (h) X_{ip}. (i) Compensated image.

does exist enough structure on the shirt of the man on the edge between trees and clouds. The $\alpha(x)$ function shown in (g) shows that the gradient is necessary on the door and on the glass door. On the sky we use both gradient and intensity. We show in figure (h) in which points of the the exhaustive flow u_e are taking account to guide the optical flow. We see that there are red points on the shirt of the man.

Figure 14.17 shows the results obtained by our model for the Walking sequence. (d) presents the estimated occlusion mask. In the video sequence the man walks in direction to the camera, in these three frames the movement of the man is very small but there is a shadow on the wall that moves. The estimations of the occlusion is noisy but preserves structure of the image. In (e) we observe the estimated optical flow. By comparing our results

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)
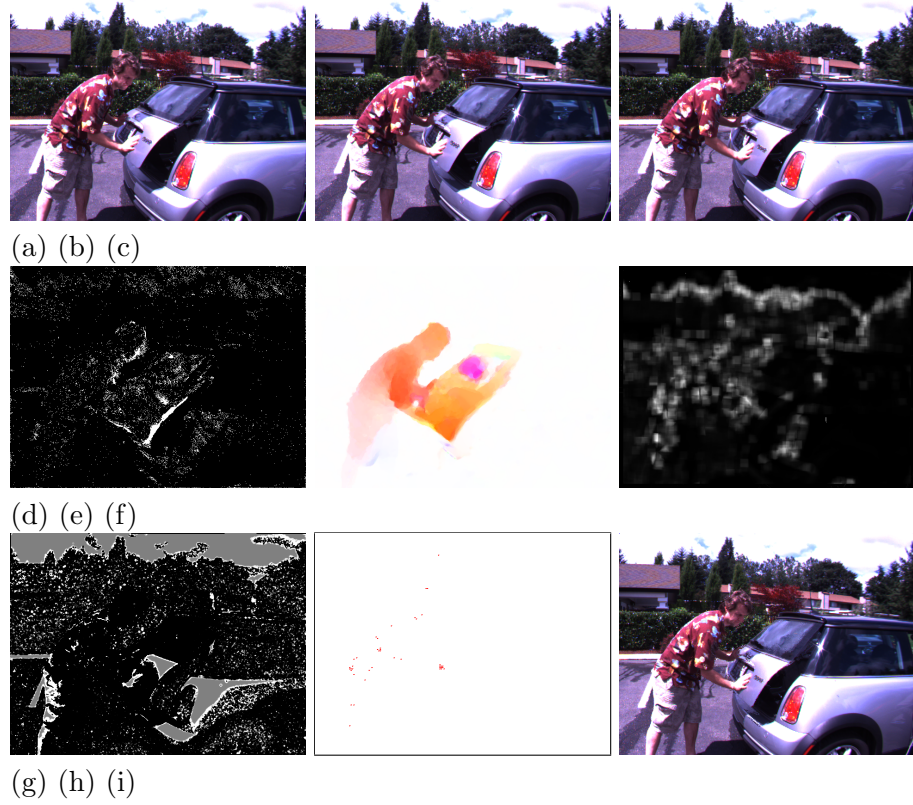
Figure 14.16: Obtained results for the MiniCooper video sequence. Frames 09, 10 and 11 are shown in (a), (b) and (c), respectively. (d): $\chi(x)$ function. (e): estimated optical flow. (f) $\lambda(x)$. (g) $\alpha(x)$. (h) X_{ip}. (i) Compensated image.

with the ones in 14.5 we observe that our model assigns zero displacement to the zone of the wall that constains shadow. In (f) we show the smaller value of the structure tensor computed in I_0. The figure shows that does exist sufficient structure on the shirt of the man and on the edge on small the table. The $\alpha(x)$ function (g) shows that the gradient is necessary almost in the whole image. We show in figure (h) in which points of the the exhaustive flow u_e are taking account to guide the optical flow. We see that there are red points on the shirt of the man.
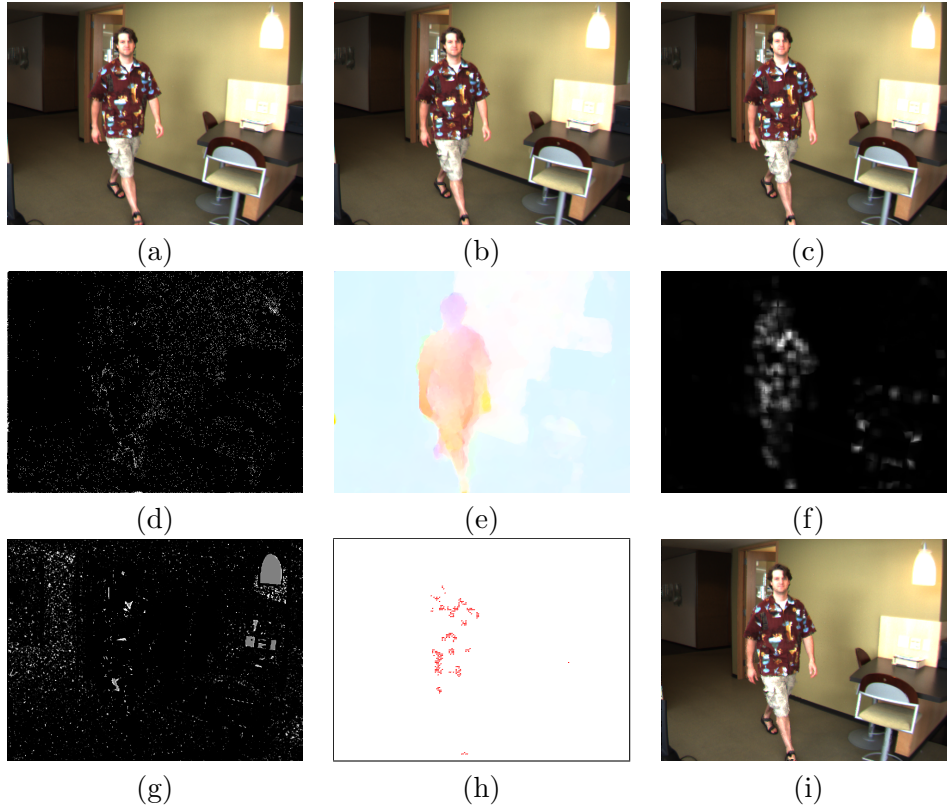
Figure 14.17: Obtained results for the Walking video sequence. Frames 09, 10 and 11 are shown in (a), (b) and (c), respectively. (d): $\chi(x)$ function. (e): estimated optical flow. (f) $\lambda(x)$. (g) $\alpha(x)$. (h) X_{ip}. (i) Compensated image.

14.6   Evaluation in MPI

The MPI-Sintel database Butler et al. (2012) presents long synthetic video sequences containing large displacements and several image degradations as blur or reflections as well as differents effects as fog or shadows. Moreover, there are two versions of the Sintel database: clean and final. The final version is claimed to be more challenging and includes all the effects previously mentioned. For our evaluation we take the final version of video sequences. Figure 14.18 displays some examples of the MPI database. There are images with large displacements, around 170 pixels for cave4, around 300 pixels for temple3. In the cave_4 sequence a girl fight with a dragon inside a cave, shown in the first and in second row of of Figure 14.18. In the first row the

girl moves her lance to attack the dragon. In second row dragon moves its jaws very fast. In the third raw we show the girl trying to cath the small dragon but a claw appears and take the small dragon. In the fourth row the girl fall on the snow. We observe large displacement and deformation of her hands.

Figure 14.18: Some challenging images of the MPI database video sequence. From top to bottom: First row: frame 10 and 11 of cave_4 video sequence. Second row: frame 45 and 46 of cave_4 video sequence. Third row: frame 30 and 31 of temple3 video sequence. Forth row: frame 06 and 07 of ambush_4 video sequence.

As a first experiment, we have computed the optical flow between frame 18 and frame 19 of the sequence market_6, by considering three frames: frame 17 which is considered to be I_{-1} in our energy model, and frame 18 which is I_1. The results are shown in Figure 14.6. The original frames17, 18 and 19 correspond to subfigures (a), (b) and (c), respectively. We observe that the occlusion mask shown in (e) is correctly estimated. Regions in the border of

the image disappear from frame18 to frame19. In (f) we show the estimated optical flow. In (g) we show the minimum eigenvalue of the structure tensor of the frame18. In (h) we show the adaptative balance term $\alpha(x)$. This show that in regions where the color constancy constrains holds, intensities are used ($\alpha(x) = 1$). In the other hand, where shadows are present, gradient is used ($\alpha(x) = 0$) (e.g., on the barrel or below the wooden wagon). Finally in (i) we show the mask where the exhaustive search is computed. In this example the End Point Error is EPE=1.7156.

We have computed the optical flow using the three frames 13, 14 15 of the sequence cave_4. The results are shown in Figure 14.21. The frames 13, 14 and 15 corresponds to subfigures (a), (b) and (c) respectively. We observe that the occlusion mask in (e) correctly estimates the occlusion due to the movement of the dragon and the occlusion due to the movement of the girl. In (f) we show the estimated optical flow. (g) displays the minimum eigenvalue of the structure tensor of frame18. In (h) we show the adaptative balance term which shows that in regions where the color constancy constrains holds intensities are used ($\alpha(x) = 1$). On the other hand, where shadows are presents gradient is used ($\alpha(x) = 0$), for examples on the back of the girl or under the wing of the dragon. Finally in (i) we show the mask where the exhaustive search is computed. In this example the end point error is EPE=2.9857.

Figure 14.22 shows the estimation of the occlusion layer obtained by our model together with the ground truth occlusion layer provided by MPI Sintel database. Our estimated occlusions are thinner.

We have also quantitatively evaluated our proposed model in the final version of the MPI database for video sequences with large displacement (sequences: ambush_2, ambush_4, ambush_5, ambush_6, market_5, market_6, cave_2, cave_4, temple_3). Actually, we divided the database in three subsets: large displacements, medium displacements and small displacements. The quatitative results are shown in Table 14.8. For large displacements we set the parameter $v_{max} = 150$, for medium displacements we set $v_{max} = 40$ and for small displacement we set $v_{max} = 1$. For large displacement we set $\theta_{lambda} = 0.50$ and $\theta_e = 0.50$, for medium and small displacement we set $\theta_\lambda = 0.98$ and $\theta_e = 0.98$. The Average End Point Error for each video sequence in the subset of the database is presented in Table 14.8.

Let us observe from Table 14.8 that the obtained average is $EPE = 18.82$ in the Large Displacement video sequences. The average EPE in all sequence reach 7.17. If we only consider the subset of frames of those sequences

Table 14.8: Summary of results. End Point Error obtained by our model in subset: large displacement, medium displacement and small displacement of MPI.

Large Displacement		Medium Displacement		Small Displacement	
Sequence	EPE	Sequence	EPE	Sequence	EPE
Ambush_2	47.94	Alley_2	0.28	Alley_1	0.33
Ambush_4	23.16	Bamboo_1	0.40	Ambush_7	2.13
Ambush_5	13.75	Bamboo_2	1.32	Bandage_1	0.60
Ambush_6	20.05	Market_2	1.00	Bandage_2	0.50
Market_5	21.71	Temple_2	1.15	Montain_1	0.84
Market_6	6.03			Shaman_2	0.28
Cave_2	26.74			Shaman_3	0.52
Cave_4	5.63			Sleeping_1	1.81
Temple_3	21.71			Sleeping_2	0.07
Average EPE	18.82	Average EPE	1.41	Average EPE	0.80

Total Average EPE	7.17

containing displacements less than 150 pixels (that is, in the range of our parameter $v_{max} = 150$ which actually imposes a bound on the ability of our method to capture the large displacement), the error drops. Table 14.9 shows the obtained subset of quantitative results. We present in Table 14.9 our results obtained in a subset of Large Displacement MPI. We also show in Table 14.9 the results obtained in this subset of data by two well-known large displacement methods, namely, DeepFlow (Weinzaepfel et al. (2013)) and Motion Detail Preserving (Xu et al. (2012)).

Table 14.9: End Point Error obtained by our model in subset of MPI with less displacement < 150 pixels.

Large Displacement	Our model	DeepFlow	MDP-Flow2
Sequence	EPE		EPE
Ambush_2	29.15	14.74	12.08
Ambush_4	12.17	14.65	15.57
Ambush_5	7.87	8.33	6.59
Ambush_6	12.91	9.93	8.47
Market_5	14.31	15.06	12.82
Market_6	3.96	6.61	5.38
Cave_2	16.98	10.08	8.48
Cave_4	3.82	4.23	3.82
Temple_3	14.67	11.90	9.01
Average EPE	8.82	10.61	9.12

(a)   (b)

(c)   (d)

(e)   (f)

(g)   (h)

(i)   (j)
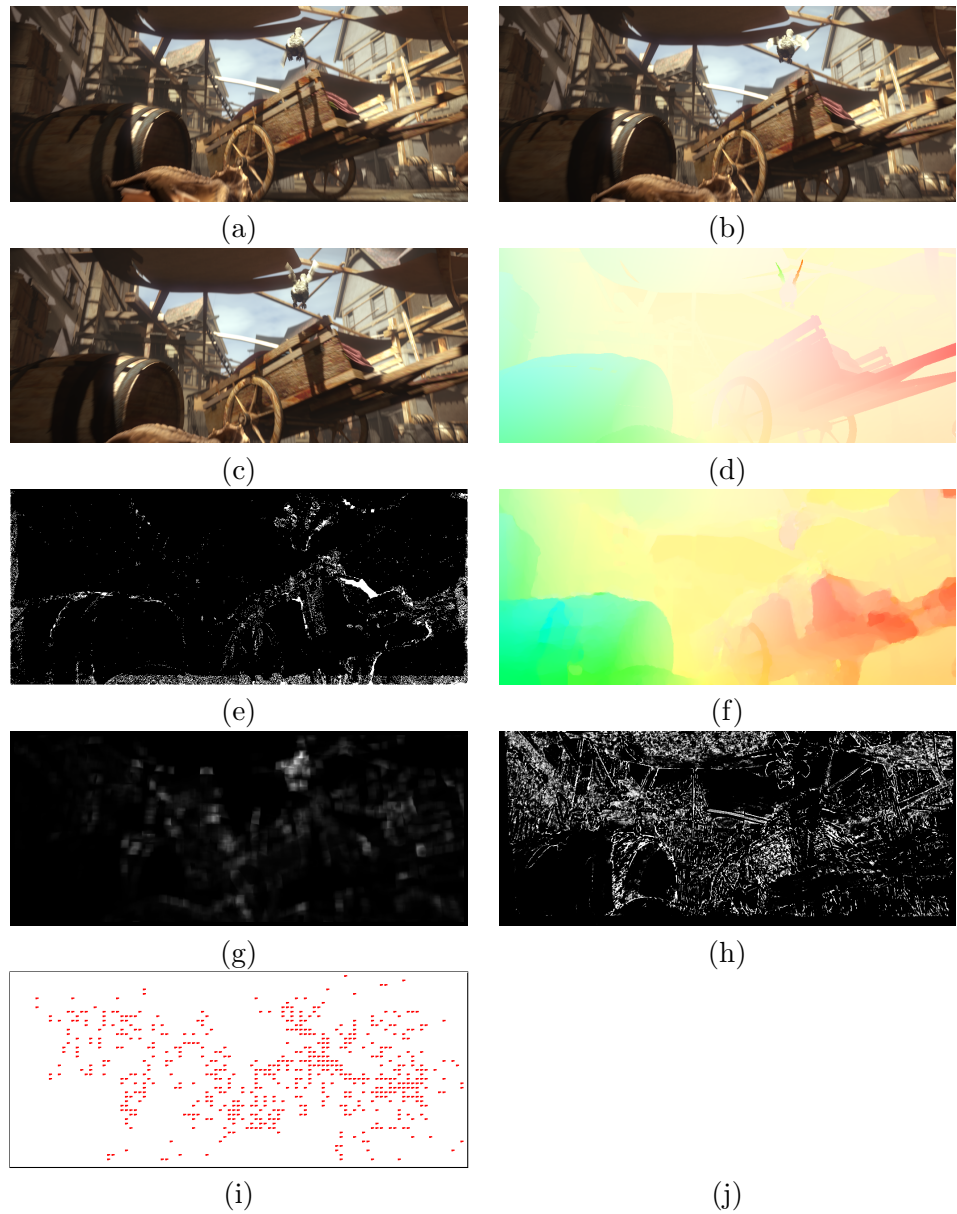
Figure 14.19: Results obtained by our method in market_6 video sequence. Frame 17, 18 and 19 are shown in (a), (b) and (c), respectively. (d) Ground truth. (e) Estimated occlusion mask $\chi(x)$. (f) Estimated optical flow. (g) minimum eigenvalue $\lambda(x)$ of the structure tensor. (h) Adaptative balance term $\alpha(x)$. (i) X_{ip} mask indicating additional matchings based on the confidence function.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)
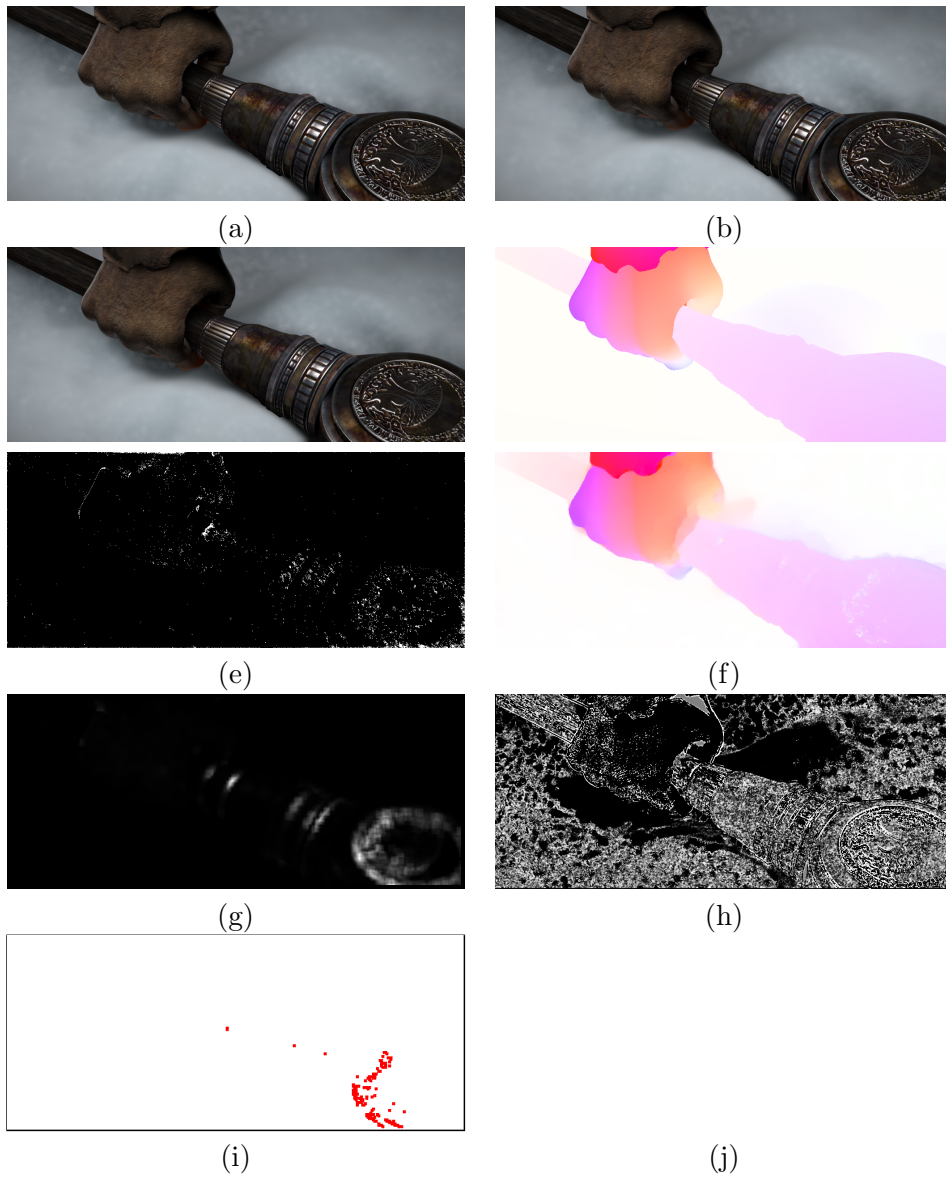
Figure 14.20: Results obtained by our method in ambush7 video sequence. (a), (b) and (c): frame 26, 27 and 28, respectively. (d) Ground truth. (e) $\chi(x)$. (f) Estimated optical flow. (g) $\lambda(x)$. (h) $\alpha(x)$. (i) X_{ip} mask where exhaustive search is computed.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)
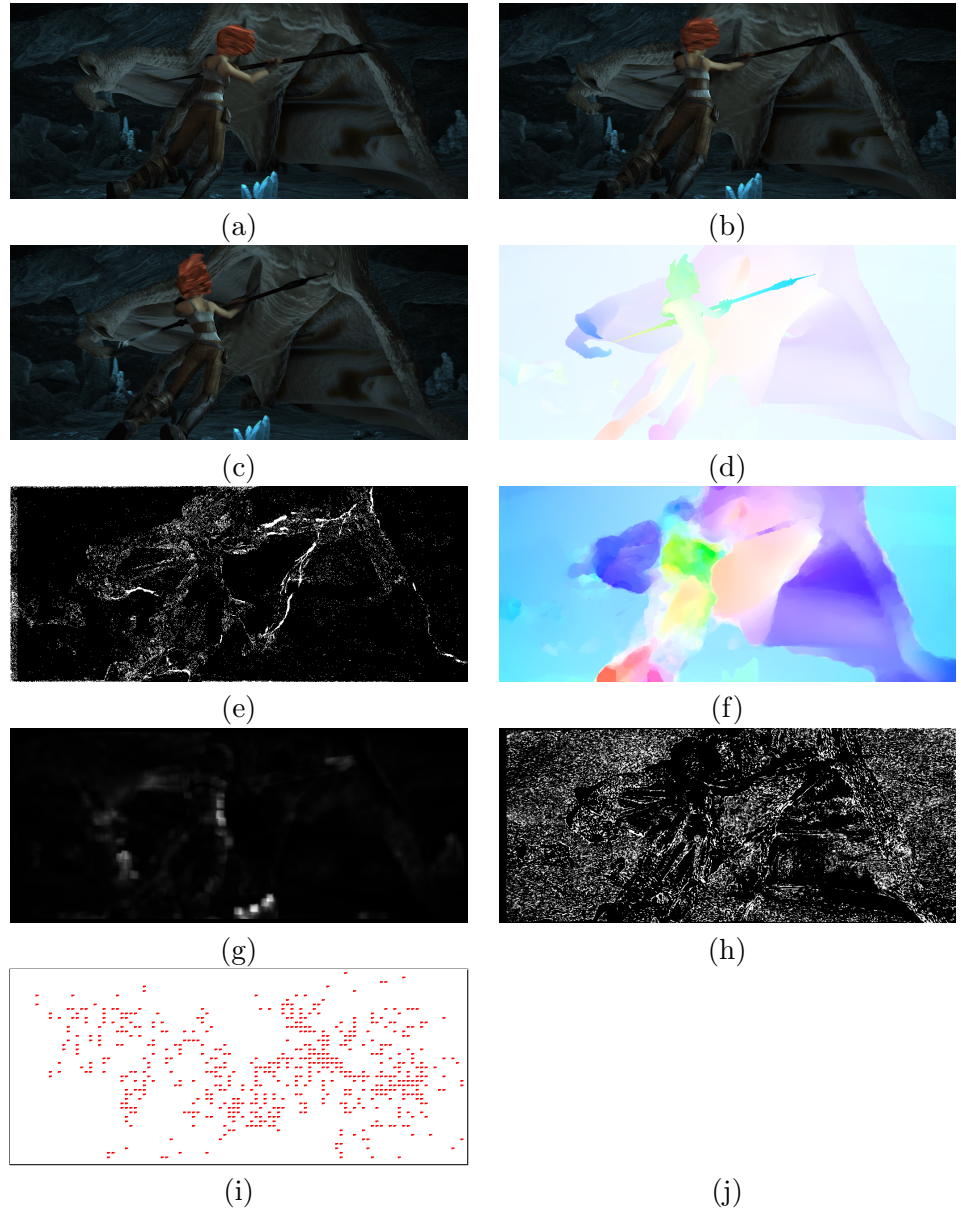
Figure 14.21: Results obtained by our method in cave4 video sequence. (a), (b) and (c): frame 13, 14 and 15, respectively. (d) Ground truth. (e) $\chi(x)$. (f) Estimated optical flow. (g) $\lambda(x)$. (h) $\alpha(x)$. (i) X_{ip} mask where exhaustive search is computed.

Figure 14.22: Occlusion layer estimation in the a video sequence of MPI. From top to bottom. First row: frame 19 and frame 20 Alley_1 video sequences. Second row: frame 21 and optical flow ground truth in video sequence Alley_1. Third row: occlusion ground truth and estimated occlusion layer.

Conclusions and future work

15.1 CONCLUSIONS

We have presented a method to estimate jointly optical flow and occlusions for realistic situations. In realistic situations occlusions, illumination changes and large displacements occurs. Traditional optical flow models may fail in realistic situations.

In this part of the thesis the first model we propose is a variational model that jointly estimate the optical flow and the occlusions layer. This model incorporate the occlusion information in its energy. Based on the divergence of the flow the information on occlusions is incorporated to the energy. Assuming that pixels occluded in the next frame are visible in the previous frame, this model considers three gray level consecutive images. The optical flow on visible pixels is computed forward. The flow on occlude pixels is computed backward.

The second model we present is a variational model that estimate jointly optical flow and occlusions. This second model is an extension of our first model to color images. This second model handles illuminations changes using a balance term between gradients and intensities. We demonstrate that the inclusion of this balance term improves the performance of the optical flow in scenarios with illumination changes, reflexions and shadows.

The third model we present is a variational model that is an extension of our second model. This model include supplementary matches given by exhaustive search in specifics locations. This supplementary marches helps to follow large displacements.

We also have presented a model to compute simultaneously forward and backward flow. The main idea is to give time consistency to the flow estimation. Preliminary results are presented.

15.2 Future work

Autosimilarity of the images, regions with lack of texture make the exhaustive matchings to fail. A robust matching method may improve the obtained results.

An implementation in a real parallel platform (for this model) is needed. State of the art GPU permits to implement near real time optical flow implementations.

Presented models depends strongly in the occlusion layer estimation. In scenarios with large displacement the estimated occlusion tends to be sparse. The forward and backward flow estimation may help to improve the estimation of occlusion layer in scenarios with large displacements.

PART IV

Conclusion of the thesis

Conclusions

We have presented a non-local gradient-based energy for disparity map interpolation, knowing a reference image. We have taken advantage of the given image both to guide the interpolation and to specify the location of depth discontinuities. We have shown with synthetic examples and two real examples the ability of our method to interpolate planes, concluding that it improves on the bilateral filter. Its Euler-Lagrange equation can be seen as a non-local fourth order equation, which can be solved as an iteration of second order Poisson equations.

The proposed functional can be used as a regularizer in a full stereo algorithm.

Our method presents some limitation in the case of textured images. The completion of the disparity will follow the texture of the reference image. Some regions of the completed disparity will stay isolated of the completion process. Another limitation appares in the case of completion of huge region of incomplete information of the diparity. The algorithm converge before the disfussion process cover the entire huge regions.

In the case we have a depth image that presents large region with lack of information, the axiomatic stated by Caselles et al. (2006) lets us construct an interpolator to complete these large empty regions.

The interpolator is computed as a ratio between two terms: a sum of function values weighted by distances and a sum of distances. This simple operation made the operator easy to implement, simple and fast.

The numerical implementation consider many very strong approximation

153

that experimentally we have shown that their are valid.

We have evaluated the Biased AMLE in differents experiments: upsampling, temporal extension. Upsampling experiments. Results show that the AMLE and Biased AMLE outperform the bilateral filter. In upsampling experiments, in the case of filtered depth images, the biased AMLE outperforms the AMLE. That show that the inclusion of the gradient in the biased AMLE helps to estimate better the depth data surface.

We have extended the biased AMLE to temporal domain. We added a new term to the distance function. This new term considers the available depth information in the next image. The use of the optical flow helps to improve the performance of the depth interpolation in video sequences.

We have compared the biased AMLE and the gradient-based non-local bilateral filter in different experiments that show their potential and differences.

We have proposed a variational model for joint computation of occlusions and optical flow in realistic scenarios. In realistic scenarios occlusions, illumination changes and large displacements occurs. The proposed model that estimate jointly occlusion and optical flow, was adapted for both gray and color images sequences is able to handle illumination changes as well as large displacements.

16.1   Contributions

In this section we outline the main contributions of the dissertation,

- Extension of the bilateral filter to gradient domain.

- Extension to the biased AMLE to time domain.

- Numerical implementation of the biased AMLE operator based on eikonal operator.

- A model to jointly estimate optical flow and occlusions which is capable to tackle illumination changes and large displacements.

List of Figures

155

List of Tables

PART V

Appendix

Proofs of Theorem Γ-limit of the gradient-based neighborhood filter

In this appendix we proof formal results of Chapter 2.1.4.

A.0.1 The local model obtained as Γ-limit of the gradient-based neighborhood filter

In order to justify the performance of the gradient-based model to produce higher order interpolations we prove that the underlying local model obtained by asymptotic rescaling of the neighborhood functions is a second order anisotropic energy that incorporates the information of the reference image. This adds further justification to our experiments that exhibit this higher order interpolation behavior. The proofs is based on the results in Bourgain et al. (2001); Ponce (2004). For the approximation of local diffusion equations by their non-local counterparts we refer to Andreu et al. (2010).

Let Ω be an open bounded set in $\mathbb{R}^N$ which represents the image domain. We assume (H)$_\Omega$: there is a constant $C_\Omega > 0$ such that given any two points $x, y \in \Omega$ there exists a curve γ connecting x to y with $L(\gamma) \leq C|x - y|$, where $L(\gamma)$ denotes the length of γ. Let $I : \Omega \to \mathbb{R}$ be a given image. We assume that $I \in W^{2,\infty}(\Omega)$. In practice this means that we have convolved a given and less regular image with a Gaussian kernel.

Let $\rho \in L^1(\mathbb{R}^N)$, $\rho \geq 0$. Let $g \in W^{1,\infty}(\mathbb{R})$ be such that

$$\min_{r \in B} g(r) \geq \alpha_B > 0 \qquad \text{for any compact subset } B \subset \mathbb{R}. \qquad \text{(A.1)}$$

Let $A \in W^{1,\infty}(\Omega \times \Omega)$, $A \geq a > 0$.

Let us consider the energy $\mathcal{P} : L^2(\Omega) \to \mathbb{R}^+$ defined by

$$\mathcal{P}(u) = \int_\Omega \int_\Omega \rho(x - y) g(I(x) - I(y)) |\nabla u(x) - \nabla u(y)|^2 A(x, y) \, dx dy, \quad \text{(A.2)}$$

if $u \in W^{1,2}(\Omega)$, and $\mathcal{P}(u) = +\infty$ if $u \in L^2(\Omega) \setminus W^{1,2}(\Omega)$.

Let $\rho_\epsilon(x) = \frac{1}{\epsilon^N} \rho(\frac{x}{\epsilon})$, $x \in \mathbb{R}^N$, $\epsilon > 0$. Let us rescale this energy as

$$\mathcal{P}_\epsilon(u) = \frac{1}{\epsilon^2} \int_\Omega \int_\Omega \rho_\epsilon(x - y) g\left(\frac{I(x) - I(y)}{\epsilon}\right) |\nabla u(x) - \nabla u(y)|^2 A(x, y) \, dx dy,$$
$$\text{(A.3)}$$

if $u \in W^{1,2}(\Omega)$. If $u \in L^2(\Omega) \setminus W^{1,2}(\Omega)$, we define $\mathcal{P}_\epsilon(u) = +\infty$.

Let

$$Q(w) = \int_{\mathbb{R}^N} \rho(z) g(\langle w, z \rangle) z \otimes z \, dz. \qquad \text{(A.4)}$$

We define

$$\mathcal{P}_0(u) = \int_\Omega \text{Trace}(D^2 u(x) Q(\nabla I(x)) D^2 u(x)) A(x, x) \, dx, \qquad \text{(A.5)}$$

if $u \in W^{2,2}(\Omega)$, and $\mathcal{P}_0(u) = +\infty$ if $u \in L^2(\Omega) \setminus W^{2,2}(\Omega)$.

Theorem A.1. *The energies $\mathcal{P}_\epsilon$ Γ-converge to $\mathcal{P}_0$ as $\epsilon \to 0+$.*

This result will be a consequence of Propositions A.2 and A.4.

Proposition A.2. *Let $u_\epsilon \in W^{1,2}(\Omega)$, $u_\epsilon \to u$ in $L^1(\Omega)$. Then*

$$\mathcal{P}_0(u) \leq \liminf_{\epsilon \to 0+} \mathcal{P}_\epsilon(u_\epsilon). \qquad \text{(A.6)}$$

Proof. The result being obviously true if the right hand side of (A.6) is $+\infty$, we may assume that $\mathcal{P}_\epsilon(u_\epsilon)$ is bounded. By $(H)_\Omega$ we have that

$$\left| \frac{I(x) - I(y)}{\epsilon} \right| \leq C_\Omega \|\nabla I\|_\infty \frac{|x - y|}{\epsilon}. \qquad \text{(A.7)}$$

Since the right hand side is bounded when $\frac{x-y}{\epsilon}$ is in the support of ρ, using (A.1) we have that

$$\int_\Omega \int_\Omega \rho_\epsilon(x-y)\frac{|\nabla u_\epsilon(x) - \nabla u_\epsilon(y)|^2}{\epsilon^2} A(x,y)\,dxdy \leq \frac{1}{\alpha}\mathcal{P}_\epsilon(u_\epsilon) \leq C, \quad \text{(A.8)}$$

for some $\alpha > 0$ and some constant $C > 0$.

Let us prove that we may assume that ρ is of compact support in $\mathbb{R}^N$. Otherwise we replace ρ by $\rho\chi_{B(0,r)}$ and we prove that

$$\mathcal{P}_0(\rho\chi_{B(0,r)}, u) \leq \liminf_{\epsilon \to 0+} \mathcal{P}_\epsilon(\rho\chi_{B(0,r)}, u_\epsilon) \leq \liminf_{\epsilon \to 0+} \mathcal{P}_\epsilon(\rho, u_\epsilon). \quad \text{(A.9)}$$

where we made explicit the dependence of the energies on the kernel $\rho\chi_{B(0,r)}$. Letting $r \to \infty$, we obtain (A.6). Thus, we assume that ρ is of compact support in $\mathbb{R}^N$. Without loss of generality assume that the support of ρ is the ball $B(0,1)$.

Step 1. Regularization of u_ϵ and preliminary inequalities. Let $\chi : \mathbb{R}^N \to \mathbb{R}$ be a smooth mollifying kernel with support in $B(0,1)$, $\chi \geq 0$, $\int_{\mathbb{R}^N} \chi(x)\,dx = 1$. Let $\chi_\delta(x) = \frac{1}{\delta^N}\chi(\frac{x}{\delta})$, $x \in \mathbb{R}^N$, $\delta > 0$. Let $\Omega_\delta := \{x \in \Omega : \text{dist}(x, \partial\Omega) \geq \delta\}$. Let $u_{\epsilon\delta} = \chi_\delta * u_\epsilon$. To avoid a cumbersome notation, let us write A instead of $A(x,y)$, A^h instead of $A(x-h, y-h)$, $\overline{A}$ instead of $A(x,x)$, $\chi^\delta_{x,y} = \chi_{\Omega_\delta}(x)\chi_{\Omega_\delta}(y)$, and

$$g_\epsilon(x,y) = g\left(\frac{I(x) - I(y)}{\epsilon}\right),$$

unless a more explicit notation is necessary. Then

$$\int_{\Omega_\delta} \int_{\Omega_\delta} \rho_\epsilon(x-y)g_\epsilon(x,y)|\nabla u_{\epsilon\delta}(x) - \nabla u_{\epsilon\delta}(y)|^2 A\,dxdy$$

$$\leq \int_{\Omega_\delta} \int_{\Omega_\delta} \int_{B(0,\delta)} \chi_\delta(h)\rho_\epsilon(x-y)g_\epsilon(x,y)|\nabla u_\epsilon(x+h) - \nabla u_\epsilon(y+h)|^2 A\,dxdydh$$

$$= \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \int_{B(0,\delta)} \chi_\delta(h)\chi^\delta_{x,y}\rho_\epsilon(x-y)g_\epsilon(x,y)|\nabla u_\epsilon(x+h) - \nabla u_\epsilon(y+h)|^2 A\,dxdydh$$

$$= \int_\Omega \int_\Omega \int_{B(0,\delta)} \chi_\delta(h)\chi^\delta_{x-h,y-h}\rho_\epsilon(x-y)g_\epsilon(x-h,y-h)|\nabla u_\epsilon(x) - \nabla u_\epsilon(y)|^2 A^h\,dxdydh$$

$$\leq \int_\Omega \int_\Omega \rho_\epsilon(x-y)g_\epsilon(x,y)|\nabla u_\epsilon(x) - \nabla u_\epsilon(y)|^2 A\,dxdy$$

$$+ \int_\Omega \int_\Omega \int_{B(0,\delta)} \chi_\delta(h)\chi^\delta_{x-h,y-h}\rho_\epsilon(x-y)Q_\epsilon(x,y,h)|\nabla u_\epsilon(x) - \nabla u_\epsilon(y)|^2\,dxdydh$$

$$:= T_1 + T_2,$$

where
$$Q_\epsilon(x, y, h) = g_\epsilon(x - h, y - h)A^h - g_\epsilon(x, y)A.$$

When $x, y \in \Omega$, $x - h, y - h \in \Omega_\delta$, the segments $[x - h, x], [y - h, y] \subset \Omega$ and, using $(H)_\Omega$, we have

$$\left| \frac{I(x - h) - I(y - h)}{\epsilon} - \frac{I(x) - I(y)}{\epsilon} \right| \leq \frac{|h|}{\epsilon} \int_0^1 |\nabla I(x - sh) - \nabla I(y - sh)| \, ds$$
$$\leq C_\Omega \|D^2 I\|_\infty \frac{|x - y|}{\epsilon} |h|.$$
$$(A.10)$$

Writing

$$g_\epsilon(x-h, y-h)A^h - g_\epsilon(x, y)A = g_\epsilon(x-h, y-h)(A^h - A) + (g_\epsilon(x-h, y-h) - g_\epsilon(x, y))A,$$

and using (A.8) and (A.10) we have

$$\frac{1}{\epsilon^2}|T_2| \leq C_1 \int_\Omega \int_\Omega \int_{B(0,\delta)} \chi_\delta(h)|h|\rho_\epsilon(x - y) \frac{|\nabla u_\epsilon(x) - \nabla u_\epsilon(y)|^2}{\epsilon^2} \, dxdydh$$
$$+ \quad C_2 \int_\Omega \int_\Omega \int_{B(0,\delta)} \chi_\delta(h)|h|\rho_\epsilon(x - y) \frac{|x - y|}{\epsilon} \frac{|\nabla u_\epsilon(x) - \nabla u_\epsilon(y)|^2}{\epsilon^2} A \, dxdydh$$
$$\leq \quad \frac{CC_1}{a}\delta + CC_2 C_\rho \delta,$$

where $C_1 = \|g\|_\infty \|A\|_{W^{1,\infty}}$, $C_2 = C_\Omega \|D^2 I\|_\infty \|\nabla g\|_\infty$, and C_ρ is a bound in the compact support of ρ and some $C > 0$.

Step 2. Let $\delta > 0$ be fixed. Let $u_\delta = \chi_\delta * u$ and $u_\delta^i = \frac{\partial u_\delta}{\partial x_i}$. We prove that

$$\frac{1}{\epsilon^2} \int_{\Omega_\delta} \int_{\Omega_\delta} \rho_\epsilon(x - y) g_\epsilon(x, y) |\nabla u_{\epsilon\delta}(x) - \nabla u_{\epsilon\delta}(y)|^2 A \, dydx$$
$$\rightarrow \quad \sum_{i=1}^N \int_{\Omega_\delta} \int_{\mathbb{R}^N} \rho(z) g(\langle \nabla I(x), z \rangle) |\langle \nabla u_\delta^i, z \rangle|^2 \overline{A} \, dzdx$$

as $\epsilon \rightarrow 0+$. First, observe that

$$\frac{1}{\epsilon^2} \int_{\Omega_\delta} \rho_\epsilon(x - y) g_\epsilon(x, y) |\nabla u_{\epsilon\delta}(x) - \nabla u_{\epsilon\delta}(y)|^2 A \, dy$$
$$\leq \quad \|D^2 u_{\epsilon\delta}\|_\infty^2 \|g\|_\infty \|A\|_\infty \int_{\Omega_\delta} \frac{|x - y|^2}{\epsilon^2} \rho_\epsilon(x - y) \, dy$$
$$\leq \quad C(\delta)$$

for any $x \in \Omega_\delta$ and $\epsilon < \delta$, where $C(\delta)$ is a constant that does not depend on ϵ.

Let $u_{\epsilon\delta}^i = \frac{\partial u_{\epsilon\delta}}{\partial x_i}$. Let $y = x - \epsilon z$, $z \in B(0,1)$. Since

$$g\left(\frac{I(x) - I(x - \epsilon z)}{\epsilon}\right) = g\left(\int_0^1 \langle \nabla I(x - s\epsilon z), z \rangle \, ds\right),$$

we have

$$\frac{1}{\epsilon^2} \int_{\Omega_\delta} \rho_\epsilon(x - y) g_\epsilon(x, y) |\nabla u_{\epsilon\delta}(x) - \nabla u_{\epsilon\delta}(y)|^2 A \, dy$$

$$= \sum_{i=1}^N \int_{\mathbb{R}^N} \chi_{\Omega_\delta}(x - \epsilon z) \rho(z) g_\epsilon(x, x - \epsilon z) \frac{|u_{\epsilon\delta}^i(x) - u_{\epsilon\delta}^i(x - \epsilon z)|^2}{\epsilon^2} A(x, x - \epsilon z) \, dz$$

$$= \sum_{i=1}^N \int_{\mathbb{R}^N} \chi_{\Omega_\delta}(x - \epsilon z) \rho(z) g\left(\int_0^1 \langle \nabla I(x - s\epsilon z), z \rangle \, ds\right)$$

$$\left|\int_0^1 \langle \nabla u_{\epsilon\delta}^i(x - s\epsilon z), z \rangle \, ds\right|^2 A(x, x - \epsilon z) \, dz$$

$$\longrightarrow \sum_{i=1}^N A(x, x) \int_{\mathbb{R}^N} \rho(z) g(\langle \nabla I(x), z \rangle) |\langle \nabla u_\delta^i, z \rangle|^2 \, dz$$

as $\epsilon \to 0+$. Then Step 2 follows by the dominated convergence theorem.

Step 3. Let us prove that

$$\sum_{i=1}^N \int_\Omega \int_{\mathbb{R}^N} \rho(z) g(\langle \nabla I(x), z \rangle) |\langle \nabla u^i(x), z \rangle|^2 \overline{A} \, dz dx$$

$$\leq \liminf_{\epsilon \to 0+} \frac{1}{\epsilon^2} \int_\Omega \int_\Omega \rho_\epsilon(x - y) g_\epsilon(x, y) |\nabla u_\epsilon(x) - \nabla u_\epsilon(y)|^2 A \, dx dy.$$

Let $\Omega' \subset\subset \Omega$. Let $\delta > 0$ be small enough so that $\Omega' \subset\subset \Omega_\delta \subset\subset \Omega$. Using the results of Step 1 and 2 and letting $\epsilon \to 0+$ we have

$$\sum_{i=1}^N \int_{\Omega'} \int_{\mathbb{R}^N} \rho(z) g(\langle \nabla I(x), z \rangle) |\langle \nabla u_\delta^i(x), z \rangle|^2 \overline{A} \, dz dx$$

$$\leq \liminf_{\epsilon \to 0+} \frac{1}{\epsilon^2} \int_\Omega \int_\Omega \rho_\epsilon(x - y) g_\epsilon(x, y) |\nabla u_\epsilon(x) - \nabla u_\epsilon(y)|^2 A \, dx dy + C(\delta),$$

where $C(\delta)$ is a constant independent of ϵ. Letting $\delta \to 0+$ and $\Omega' \uparrow \Omega$, Step 3 follows.

Step 4. Conclusion. Since

$$\sum_{i=1}^{N} \rho(z)g(\langle \nabla I(x), z \rangle)|\langle \nabla u^i(x), z \rangle|^2$$

$$= \mathrm{Trace}\left(\rho(z)g(\langle \nabla I(x), z \rangle)(z \otimes z)\sum_{i=1}^{N}\left(\nabla u^i(x) \otimes \nabla u^i(x)\right)\right),$$

then

$$\int_{\mathbb{R}^N}\sum_{i=1}^{N}\rho(z)g(\langle \nabla I(x), z\rangle)|\langle \nabla u^i(x), z\rangle|^2 = \mathrm{Trace}\left(Q(\nabla I(x))\sum_{i=1}^{N}\left(\nabla u^i(x) \otimes \nabla u^i(x)\right)\right)$$

$$= \mathrm{Trace}\left(D^2u(x)Q(\nabla I(x))D^2u(x)\right).$$

Then (A.6) follows from this and Step 3.

$$\square$$

Let us recall the following simple result.Bourgain et al. (2001)

Lemma A.3. *Assume that Ω is a bounded domain in $\mathbb{R}^N$ with Lipschitz boundary. Let $w \in W^{1,p}(\Omega)$, $1 \le p < \infty$ and let $\rho \in L^1(\mathbb{R}^N)$, $\rho \ge 0$. Then*

$$\int_{\Omega}\int_{\Omega}\rho(x-y)\frac{|w(x)-w(y)|^p}{|x-y|^p}\,dxdy \le C\|w\|_{W^{1,p}}\|\rho\|_1. \tag{A.11}$$

Proposition A.4. *Let $u \in W^{2,2}(\Omega)$. Then*

$$\mathcal{P}_0(u) = \lim_{\epsilon \to 0+}\mathcal{P}_\epsilon(u). \tag{A.12}$$

Proof. Since

$$\mathrm{Trace}\left(D^2u(x)Q(\nabla I(x))D^2u(x)\right) = \sum_{i=1}^{N}\langle Q(\nabla I(x))\nabla u^i(x), \nabla u^i(x)\rangle,$$

it suffices to prove that

$$\lim_{\epsilon \to 0+}\int_{\Omega}\int_{\Omega}\rho_\epsilon(x-y)g\left(\frac{I(x)-I(y)}{\epsilon}\right)\frac{|u^i(x)-u^i(y)|^2}{\epsilon^2}A(x,y)\,dxdy$$

$$= \int_{\Omega}\langle Q(\nabla I(x))\nabla u^i(x), \nabla u^i(x)\rangle A(x,x)\,dx$$

$$\tag{A.13}$$

for all $i = 1, \ldots, N$.

First we observe that it suffices to assume that $u \in C^2(\Omega)$. Let us write
$v = u^i$. Let $w \in W^{1,2}(\Omega)$. Let

$$
V_\epsilon = \sqrt{\rho_\epsilon(x - y)} \sqrt{g\left(\frac{I(x) - I(y)}{\epsilon}\right)} \frac{|v(x) - v(y)|}{\epsilon} \sqrt{A(x, y)},
$$

$$
W_\epsilon = \sqrt{\rho_\epsilon(x - y)} \sqrt{g\left(\frac{I(x) - I(y)}{\epsilon}\right)} \frac{|w(x) - w(y)|}{\epsilon} \sqrt{A(x, y)}.
$$

Then

$$
\left| \|V_\epsilon\|_2 - \|W_\epsilon\|_2 \right|^2 \le \|V_\epsilon - W_\epsilon\|_2^2
$$
$$
= \int_\Omega \int_\Omega \rho_\epsilon(x - y) g\left(\frac{I(x) - I(y)}{\epsilon}\right) \frac{\left| |v(x) - v(y)| - |w(x) - w(y)| \right|^2}{\epsilon^2} A(x, y)\, dxdy
$$
$$
\le \int_\Omega \int_\Omega \rho_\epsilon(x - y) g\left(\frac{I(x) - I(y)}{\epsilon}\right) \frac{\left| (v(x) - v(y)) - (w(x) - w(y)) \right|^2}{\epsilon^2} A(x, y)\, dxdy
$$

Since ρ has compact support $\rho_\epsilon(x - y) \ne 0$ if and only if $\frac{|x - y|}{\epsilon} \le C$ for
some constant $C > 0$. In that case $\frac{1}{\epsilon} \le \frac{C}{|x - y|}$. We can continue the above
inequalities

$$
\le C \int_\Omega \int_\Omega \rho_\epsilon(x - y) g\left(\frac{I(x) - I(y)}{\epsilon}\right) \frac{\left| (v(x) - v(y)) - (w(x) - w(y)) \right|^2}{|x - y|^2} A(x, y)\, dxdy
$$
$$
\le C\|g\|_\infty \|\rho\|_1 \|A\|_\infty \|v - w\|_{W^{1,2}(\Omega)},
$$

where the last inequality follows from Lemma A.3. Thus, by density we
may assume that $u^i \in C^1(\Omega)$, i.e., $u \in C^2(\Omega)$. Then proceeding as in Step
2 of Proposition A.2 we obtain (A.13). $\qquad\square$

Proposition A.5. *Let us consider* $\rho(z) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{|z|^2}{2\sigma^2}}$, $\sigma > 0$, $z \in \mathbb{R}^N$,
$g(r) = e^{-\frac{r^2}{2h^2}}$, $r \in \mathbb{R}$, $h > 0$. *Let* $Q(w)$ *be given by* (A.4). *Let* $B = \frac{1}{\sigma^2}I + \frac{1}{h^2}w \otimes w$. *Then*

$$
\begin{aligned}
Q(w) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} e^{-\frac{\langle Bz, z \rangle}{2}} z \otimes z\, dz \\
&= \frac{1}{(2\pi\sigma^2)^{N/2}} (\det B)^{-1/2} \int_{\mathbb{R}^N} e^{-\frac{\langle \bar{z}, \bar{z} \rangle}{2}} B^{-1/2}\bar{z} \otimes B^{-1/2}\bar{z}\, d\bar{z}.
\end{aligned}
$$

Since

$$\det B = \frac{1}{\sigma^{2N}}\left(1 + \frac{\sigma^2}{h^2}|w|^2\right),$$

and

$$\frac{1}{(2\pi)^{N/2}}\int_{\mathbb{R}^N} e^{-\frac{\langle \bar{z},\bar{z}\rangle}{2}}\,\bar{z}\otimes\bar{z}\,d\bar{z} = I,$$

we have

$$Q(w) = \sigma^2\left(1 + \frac{\sigma^2}{h^2}|w|^2\right)^{-1/2}\left(I + \frac{\sigma^2}{h^2}w\otimes w\right)^{-1}.$$

Thus

$$Q(\nabla I(x)) = \sigma^2\left(1 + \frac{\sigma^2}{h^2}|\nabla I(x)|^2\right)^{-1/2}\left(I + \frac{\sigma^2}{h^2}\nabla I(x)\otimes\nabla I(x)\right)^{-1}$$

is an anisotropic tensor.

Proposition A.6. *In a similar way, if*

$$\mathcal{B}_\epsilon(u) = \frac{1}{\epsilon^2}\int_\Omega\int_\Omega \rho_\epsilon(x-y)g\left(\frac{I(x)-I(y)}{\epsilon}\right)|u(x)-u(y)|^2 A(x,y)\,dxdy, \quad u\in L^2(\Omega),$$
(A.14)

and we define

$$\mathcal{B}_0(u) = \int_\Omega\langle Q(\nabla I(x))(\nabla u(x)),\nabla u(x)\rangle A(x,x)\,dx, \qquad (A.15)$$

if $u\in W^{1,2}(\Omega)$, and $\mathcal{B}_0(u) = +\infty$ if $u\in L^2(\Omega)\setminus W^{1,2}(\Omega)$, we have

Theorem A.7. *As $\epsilon\to 0+$, the energies $\mathcal{B}_\epsilon$ Γ-converge to the energy $\mathcal{B}_0$.*

Proposition A.8. *One can also compute the Γ-limit in $L^2(\Omega)$ of bilateral filter energies in the case of a faster rescaling*

$$\mathcal{Q}_\epsilon^a(u) = \frac{1}{\epsilon^2}\int_\Omega\int_\Omega \frac{\rho_\epsilon(x-y)}{\max(C_\epsilon(x),C_\epsilon(y))}g\left(\frac{I(x)-I(y)}{\epsilon^{1+\alpha}}\right)|u(x)-u(y)|^2 A(x,y)\,dxdy,$$
(A.16)

where $u\in L^2(\Omega)$, $\alpha > 0$, and

$$C_\epsilon(x) = \int_\Omega \rho_\epsilon(x-y)g\left(\frac{I(x)-I(y)}{\epsilon^{1+\alpha}}\right)\,dy.$$

In that case the matrix $Q(\nabla I(x))$ is replaced by $Q^a(\nabla I(x)) = cP_x$ where $P_x = I - \frac{\nabla I(x)\otimes\nabla I(x)}{|\nabla I(x)|^2}$, if $\nabla I(x)\neq 0$, and $P_x = I$ if $\nabla I(x) = 0$. We assume that the set of points $x\in\Omega$ where $\nabla I(x) = 0$ is a null set. We notice that it is also not difficult to compute the Γ-limit in $W^{1,2}(\Omega)$ in the case of gradient energies with a faster rescaling although this result is not fully satisfactory.

Bibliography

A. Ayvaci, M. Raptis, and S. Soatto. Sparse occlusion detection with optical flow. *International Journal of Computer Vision*, 97(3):322–338, 2012. 78

A. Almansa, F. Cao, Y. Gousseau, and B. Roug. Interpolation of digital elevation models using amle and related methods. *IEEE Transaction on Geoience and remote sensing*, 40(2):314–325, 2002. xi, xvii, 24, 45

L. Alvarez, R. Deriche, T. Papadopoulo, and J. Sanchez. Symmetrical dense optical flow estimation with occlusions detection. *International Journal of Computer Vision*, 75(3):371–385, 2007. 77

F. Andreu, J.M. Mazon, J. D. Rossi, and J. Toledo. Nonlocal diffusion problems. *Mathematical Surveys and Monographs*, 2010. 167

P. Arias, V. Caselles, and G. Sapiro. A variational framework for non-local image inpainting. In *EMMCVPR, Lecture Notes in Computer Science, (Springer Verlag)*, pages 345–358, 2009. 4

P. Arias, G. Facciolo, V. Caselles, and G. Sapiro. A variational framework for exemplarbased image inpainting. *International Journal of Computer Vision*, pages 319–347, 2011. 4

S.N. Armstrong, C. K. Smart, and S. J. Somersille. An infinity laplace equation with gradient term and mixed boundary conditions. *arXiv preprint*, 2009. xvi, 23, 25, 36

G. Aronsson. Extension of functions satisfyng lipschitz conditions. *Aktiv fr Mathematik*, 6(6):551–561, 1967. xii, 24, 36

G. Aronsson. On the partial differential equation $u_x^2 u_{xx} + 2u_x u_y u_{xy} + u_y^2 u_{yy} = 0$. *Aktiv fuer Mathematik*, 7(5):395–425, 1968. xii, 24, 36

G. Aronsson, M. G. Grandall, and P. Juutinen. A tour of the theory of

175

absolutely minimizing functions. *Bullentin of the American mathematical society*, 41(4):439–506., 2004. 24, 36

S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal on Computer Vision*, 92(1):1–31, 2011. xviii, 51, 115

I. Galic J. Weickert M. Welk A. Bruhn A. Belyaev and H-P. Seidel. Toward pde-based image compression. *x y j i*, vol. 14(no. 3):271–284, 2005. 25, 37

M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer vision and image understanding*, 63(1):75–104, 1996. xii, 73, 75

J. Bourgain, H. Brezis, and P. Mironescu. Another look at sobolev spaces, in optimal control and partial differential equation. In *Conference in honor to A. Bensoussan 60th birthday*, pages 439–455, 2001. 4, 167, 172

A. Braides. *Gamma-convergence for Beginners.* Oxford University Press, USA, 2002. 11

T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision (ECCV)*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36, 2004a. xii, 73, 75

T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision*, volume 4, pages 25–36, 2004b. xxi, xxii, 79, 83, 84, 91, 98

T. Brox, C. Bregler, and J. Malik. Large displacement optical flow. In *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2009. 79

A. Bruhn and J. Weickert. *A confidence measure for variational optic flow methods.* Springer Netherlands, 2006. 81

A. Buades, B. Coll, and J.M. Morel. Neighborhood filters and pdes. *Numerische Mathematik*, 105:1–34, 2006. 4, 8

D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In A. Fitzgibbon et al. (Eds.), editor, *European Conf. on Computer Vision (ECCV)*, Part IV, LNCS 7577, pages 611–625. Springer-Verlag, October 2012. 80, 96, 115, 140

C. Kondermann, R. Mester, and C. Garbe. A statistical confidence measure for optical flow. In *Computer Vision-ECCV 2008*, pages 290–301, 2008. 81

F. Cao, Y. Gousseau, S. Masnou, and P. Perez. Geometrically guided exemplar-based inpainting. *SIAM Journal on Imaging Sciences*, vol. 4: 1143 – 1179, 2011. 15

V. Caselles, J.M. Morel, and C. Sbert. An axiomatic approach to image interpolation. *IEEE Transaction on image Processing*, 7(2):376–386, 1998. xvii, 23, 24, 25, 27, 28, 36

V. Caselles, L. Igual, and O. Sander. An axiomatic approach to scalar data interpolation on surfaces. *Numerische Mathematik*, 102(3):383–411, 2006. xi, xvi, xvii, 23, 24, 25, 26, 27, 28, 29, 32, 35, 67, 153

A. Chambolle. An algorithm for total variation minimization and applications. *Mathematical Imaging and Vision*, 20(1):89–97, 2004. 87

A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011. 88

Z. Chen, H. Jin, Z. Lin, S. Cohen, and Y. Wu. Large displacement optical flow from nearest neighbor fields. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 2443–2450, 2013. 73, 75

T. Corpetti, E. Mémin, and P. Pérez. Dense estimation of fluid flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):365–380, March 2002. 78

M. Crandall. A visit with infinite-laplacian equation. *Calculus of varoations and nonlinear partial differential equations*, pages 75–122., 2008. 24, 36

M.G Crandall, L.C. Evans, and R.F. Gariepy. Optimal lipschitz extensions and the inifinity laplacian. *Calculus of Variations and Partila Differential Equations*, 13(2):123–139, 2001. 24, 36

D. Fortun, P. Bouthemy, and C. Kervrann. Aggegation of local parametric candidates with exemplar-based occlusion hangling for optical flow. *Computer Vision and Image Undestanding (Accepted)*, 2015. 80

A. Desolneux, L. Moisan, and J.M. Morel. Edge detection by helmholtz principle. *Journal of Mathematical Imaging and Vision*, vol. 14(no. 3): 271–284, 2001. 15

J. Digne, Dimicolim M., N. Sabater, and P. Salembier. *Neighborhood filters and the recovery of 3d information.* Springer, 2010. xi, 4, 5, 8, 13

I. Galic, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, and H-P. Seidel. Image compression with anisotropic diffusion. *preprint*, 2015. 25, 37

I. Gómez and J.D. Rossi. Tug-of-war games and the infinity laplacian with

spatial dependence. *Communications on Pure and Applied Analysis*, 5: 1959–1983., 2013. 24, 36

D. Herrera, J. Kannala, and J. Heikkila. Accurate and practical calibration of a depth and color camera pair. In *Conf. Comp. Anal. of Images and Patterns. Springer*, pages 437–445, 2011. 16

B. K. Horn and B. H. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981. xii, 73, 75

L. Igual. *Image segmentation and compression using the tree of shapes of and image. Motion Estimation.* PhD thesis, Universitat Pompeu Fabra, 2006. 76

S. Ince and J. Konrad. Occlusion-aware optical flow estimation. *IEEE Trans. Image Processing*, 17(8):1443–1451, 2008. 77

R. Jensen. Uniqueness of lipschitz extensions: minimizing the sup norm of the gradient. *Archive for Rational Mechanics and Analysis*, 123(1): 51–74., 1993. 24, 36

P. Juutinen. *Minimization problems for Lipscitz functions via viscosity solutions.* University of Jyvaskyla, 1998. 24, 36

F. Kendoul, I. Fantoni, and Nanomia. K. Optical flow-based vision system for autonomous 3d localization and control of small aerial vehicles. *Robotics and Autonomous Systems*, 50:177–194, 2005. xii, 71

P. Krähenbühl and V. Koltun. Efficient nonlocal regularization for optical flow. In *European Conference on Computer Vision (ECCV)*, pages 356–369. Springer, 2012. 73, 75

J. Kybic and C. Nieuwenhuis. Bootstrap optical flow confidence and uncertainty measure. *Computer Vision and Image Understanding*, 115(10): 1449–1462, 2011. 81

V. Lazcano, P. Arias, and V. Caselles. A gradient based neighborhood filter for depth interpolation. In *IEEE International Conference on Image Processing(ICIP), Orlando, Florida, USA*, pages 873–876, 2012. 5

J. J. Manfredi, A. M. Oberman, and A.P. Sviridov. Nonlinear elliptic partial differential equation and p-harmonic functions on graphs. *preprint*, 2015. xvii, 26, 43

J.J Manfredi, A.M. Oberman, and A.P. Sviridov. Nonlinear elliptic partial differential equation and p-harmonic functions on graphs. *preprint*, 2012.

E. Meinhardt-Llopis, J. Sanchez, and D. Kondermann. Horn-schunck optical flow with multiscale s trategy. http://www.ipol.im/pub/art/2013/20/article_lr.pdf, 2013. xxii, 72, 84, 99

E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing*, 7(5):703–719, May 1998. xxii, 79, 84, 98

T. Müller, C. Rabe, J. Rannacher, U. Franke, and R. Mester. Illumination-robust dense optical flow using census signatures. In *Proceedings of the DAGM Conference on Pattern Recognition*, pages 236–245. Springer, 2011. 75

H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 6(5): 565–593, 1986. 73, 75

P. Ndjiki-Nya, M. Koeppel, D. Doshkov, H. Lakshman, P. Merkle, K. Mueller, and T. Wiegand. Depth image based rendering with advanced texture synthesis. In *ICME 2010*, pages 424–429, 2010. xi, xvii, 24

O. M. Aodha, A. Humayun, M. Pollefeys, and G. J. Brostow. Learning a confidence measure for optical flow. *IEEE Transaction on Pattern Analysis and Machine Inteligence*, 3(5):1107–1120, May 2013. 81

A. M. Oberman. A convergent difference scheme for the infinity laplacian: construction of absolutely minimizing lipschitz extensions. *Mathematics of Computation*, 74(251):1217–1230, 2005a.

A. M. Oberman. A convergent difference scheme for the infinity laplacian: construction of absolutely minimizing lipschitz extensions. *Mathematics of Computation*, 74(251):1217–1230, 2005b. xvii, xviii, 26, 43, 45

I. Patras, E. A. Hendriks, and R. L. Lagendijk. Probabilistic confidence measures for block matching motion estimation. *IEEE Transaction on Circuits and Systems for Video Technology*, 17(8):988–995, August 2007. 81

Y. Peres, G. Pete, and S. Somersille. Biased tuf-of-war, the biased infinity laplacian , and comparison with exponenetial cones. *Calculus of Variations and Partial Differential Equations*, 38(3):541–564., 2010. xvi, 23, 25, 36

P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on PAMI*, 12:629–39, 1990. xi, 8

A. C. Ponce. A new approach to sobolev spaces and connections to-convergence. *Calculas of Variations and Partial Diffrenectial Equations*, 19:229–255, 2004. 167

R P.Palomares, G. Haro, and C. Ballester. *A Rotation-Invariant Regular-*

ization Term for Optical Flow Related Problems, volume 9007 of *Lectures Notes in Computer Science*, pages 304–319. Springer, Proceedings of the ACCV'14, Singapore, 2015. 73, 75

R. Kennedy and C. J. Taylor. Optical flow with geometric occlusion estimation and fusion of multiple frames. *Energy Minimization Methods in Computer Vision and Pattern Recognition*, 8932:364–377, 2015. 80

R. Ranftl, K. Bredies, and T. Pock. Non-local total generalized variation for optical flow estimation. In *Computer Vision–ECCV 2014*, pages 439–454. Springer, 2014. 73, 75

F. Ruffier and N. Franceschini. Optical flow regulation: the key to aircraft automatic guidance. *Robotics and Autonomous Systems*, 50:177–194, 2005. xii, 71

J. Sánchez, A. Salgado, and N. Monzón. Preserving accurate motion contours with reliable parameter selection. In *Image Processing (ICIP), IEEE International Conference on*, pages 209–213, 2014. 73, 75

P. Sand and S. Teller. Particle video: Long-range motion estimation using point trajectories. *International Journal of Computer Vision*, 80(1):72–91, 2008a. 78

P. Sand and S. Teller. Particle video: Long-range motion estimation using point trajectory. *International Journal on Computer Vision*, 80(1):72–91, 2008b. 85

S. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, vol. 47:7–42, 2002. 15

F. Stein. Efficient computation of optical flow using the census transform. In *DGAM*, pages 79–86. Springer, 2004. 75

F. Steinbruecker and T. Pock. Large displacement optical flow computation without warping. In *International Conference on Computer Vision*, pages 1609–1614, 2009. 79, 99

C. Stiller and J. Konrad. Estimating motion in image sequences. *IEEE Signal Processing Magazine*, 16(4):70–91, 1999. xii, 71

M. Stoll, S. Volz, and A. Bruhn. Adaptive integration of features matches into variational optical flow methods. In *Proc. of the Asian Conference in Computer Vision (ACCV)*, pages 1–14, 2012. xxi, 84, 99, 100, 105

E. Strekalovskiy, A. Chambolle, and D. Cremers. Convex relaxation of vectorial problems with coupled regularization. *SIAM J. Imaging Sciences*, 7(1):294–336, 2014. 73, 75

D. Sun, S. Roth, and M. J Black. Secrets of optical flow estimation and their principles. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2432–2439, 2010a. xii, 73, 75, 78

D. Sun, E. B. Sudderth, and M. J. Black. Layered image motion with explicit occlusions, temporal consistency, and depth ordering. In *Advances in Neural Information Processing Systems*, pages 2226–2234, 2010b. 77

D. Sun, S. Roth, and M. J. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision*, 106(2):115–137, 2014. 73, 75

W. B. Thompson, K. M. Mutch, and V. A. Berzins. Dynamic occlusion analysis in optical flow fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(4):374–383, 1985. 78

R. Timofte and L. Van Gool. Sparseflow: Sparse matching for small to large displacement optical flow. In *IEEE Winter Conference On Applications of Computer Vision (WACV)*, pages 1100–1106, 2015. 80, 99

C. Tomasi and R. Manduchi. Bilateral filter for gray and color images. In *Sixth International Conference on Computer Vision (IEEE Press)*, pages 839–846, 1998. 3, 4, 5, 8

J. Y. A. Wang and E. H. Adelson. Representing moving images with layers. *IEEE Transactions on Image Processing*, 3(5):625–638, 1994. 77

P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid. Deepflow: Large displacement optical flow with deep matching. In *IEEE International Conference on Computer Vision, Sydney, Australia*, pages 1385–1392, 2013. 79, 99, 143

M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof. Anisotropic huber-l1 optical flow. In *BMVC*, volume 1, page 3, 2009. 73, 75

M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2464–2471. IEEE, 2010. 73, 75

J. J. Xiao, H. Cheng, H. S. Sawhney, C. Rao, and M. Isnardi. Bilateral filtering-based optical flow estimation with occlusion detection. In *ECCV*, pages I: 211–224, 2006. 77, 78

L. Xu., J. Jia, and Y. Matsushita. Motion detail preserving optical flow. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 78, 79, 89, 91, 93, 99

L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions*

on, 34(9):1744–1757, 2012. 73, 75, 143

Q. Yang, R. Yang, J. Davis, and D. Nister. Spatial-depth super resolution for range images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. xi, 4, 5, 8, 13

L. P. Yaroslavsky. *Digital picture processing. An introduction.* Springer Verlag, 1985. 5, 8

J. Yin and J. R. Cooperstock. Improving depth maps by nonlinear diffusion. In *Proc. of the 12th Int. Conf. on Computer Graphics Visualization and Computer Vision (WSCG)*, pages 100–101, 2004. xi, 5

K.J. Yoon and I.S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis andMachine Intelligence*, vol. 28:650–656, 2006. xi, 4, 5, 8

C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime tv-l1 optical flow. In *Proceedings of the 29th DAGM Conference on Pattern Recognition*, pages 214–223, Berlin, Heidelberg, 2007. Springer-Verlag. xii, 73, 75, 85, 87

C. Zack, T. Pock, and H. Bischof. A duality based approach for realtime tv-l1 optical flow. In *29th DAGM Symposium on Pattern Recognition2007*, pages 214–223, 2007. 76

H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388, 2011. 73, 75