2016

Sparse variational regularization for visual motion estimation

Muhammad Wasim Nawaz
University of Wollongong, mwn831@uow.edu.au

Sparse Variational Regularization for Visual Motion Estimation

A thesis submitted in partial fulfillment of the requirements for the award of the degree

Doctor of Philosophy

from

The University of Wollongong

by

Muhammad Wasim Nawaz
B. Sc. Electrical Engineering

School of Electrical, Computer and Telecommunications Engineering

February 2016

Contents

List of Figures

List of Tables

Notations and Acronyms

Notations

x	a scalar variable
$\mathbf{x}$	a vector variable
X	a matrix variable
$\lvert x \rvert$	absolute value of a scalar x
$\lvert \mathbf{x} \rvert$	element-wise absolute value of $\mathbf{x}$
$\lvert\lvert \mathbf{x} \rvert\rvert_p$	ℓ_p norm of $\mathbf{x}$
$\rho(x)$	robust penalty function for variable x
$(.)^{\mathrm{T}}$	transpose operator
$(.)_x$	horizontal derivative
$(.)_y$	vertical derivative
$(.)_t$	temporal derivative
α	a positive weight used for variational problems
η	noise variable
λ	Lagrange multiplier
ϵ	Huber norm parameter
∇	gradient operator
$\boldsymbol{\nabla}$	sparse difference matrix to calculate gradient of a vector

Δ	Laplacian operator
A	a matrix made up of first order spatial intensity derivatives
B	a matrix made up of second order spatial intensity derivatives
C	a matrix made up of spatial intensity derivatives and identity matrices
c	additive intensity change parameter
d	multiplicative intensity change parameter
$\mathfrak{D}$	domain of a video frame
$\mathbf{F}$	a multi-channel image
$\mathbf{f}$	lexicographically vectorized $\mathbf{F}$
$\mathbf{g}$	lexicographically vectorized degraded $\mathbf{F}$
I	image intensity
$\boldsymbol{I}$	vectorized image intensity in a lexicographical manner
$\mathcal{I}_n$	$n \times n$ identity matrix
L	Lipschitz constant
m	number of measurements taken in a video frame
n	number of pixels or dimensionality of a video frame
$\mathbf{q}$	joint optical flow, occlusion and intensity change parameters vector
$\mathfrak{R}$	a region or neighborhood in a video frame
s	cardinal number denoting the sparsity of a signal
t	time variable
$\mathbf{y}_1$	temporal intensity derivative vector
$\mathbf{y}_2$	spatiotemporal intensity derivative vector

Acronyms

3D	three-dimensional
FFT	fast Fourier transform
GCA	gradient constancy assumption
GDIM	generalized dynamic image model
HOTV	higher order total variation
HSV	hue, saturation and value color space
HVD	proposed regularizer that uses horizontal, vertical and diagonal derivatives
i.i.d.	independent and identically distributed
MLE	maximum likelihood estimation
MEPE	mean end point error
MSE	mean square error
NESTA	Nesterov's algorithm
NCC	normalized cross correlation
OF	optical flow
OFC	optical flow constraint
PR	precision-recall curve
RGB	red, green and blue color space
SR	sparse representation
SAD	sum of absolute differences
SSD	sum of squared differences
TGV	total generalized variation
TV	total variation

Abstract

The computation of visual motion is a key component in numerous computer vision tasks such as object detection, visual object tracking and activity recognition. Despite extensive research effort, efficient handling of motion discontinuities, occlusions and illumination changes still remains elusive in visual motion estimation. The work presented in this thesis utilizes variational methods to handle the aforementioned problems because these methods allow the integration of various mathematical concepts into a single energy minimization framework. This thesis applies the concepts from signal sparsity to the variational regularization for visual motion estimation. The regularization is designed in such a way that it handles motion discontinuities and can detect object occlusions.

The first part of this thesis estimates visual motion between two consecutive video frames by mapping co-visible scene points. This mapping is commonly known as *optical flow*. The accurate handling of motion discontinuities is essential for a better visual perception and quality of the estimated optical flow. To this end, the discontinuity preserving and sparsity promoting robust ℓ_1 norm is investigated in the context of variational regularization. A sparse variational regularizer is proposed which exploits the sparsity of the gradient field of optical flow. The signal sparsity enables us to recover a signal from a few measurements; thus, by utilizing this regularizer, the problem of recovering optical flow from a small set of linear measurements is also investigated. It is found that exploiting the sparsity of partial derivatives of optical flow results in a decrease in the computational complexity and memory requirements.

The second part of the thesis investigates the use of optical flow for the detection of the occluded video regions under varying illumination of the scene. Since the occluded video regions are usually small in any frame of a video sequence taken at a sufficiently high temporal rate, a sparsity constraint is imposed on these regions. Moreover, illumination changes are explicitly modeled to counter for geometric and photometric brightness variations of the scene. A generalized dynamic image model (GDIM) is used to construct the data term; occlusion is then detected as the sparse residual of the data term. Brightness change parameters of the GDIM are also estimated by imposing a smoothness constraint over them. A modification of this method is also presented which does not require the estimation of brightness change parameters under certain conditions. Experimental results demonstrate that the presented optical flow estimation method is able to accurately detect occluded video regions under varying illumination of the scene.

Acknowledgments

In the name of God, the Most Gracious, the Most Merciful. I thank Him for allowing me to undertake PhD studies, and bestowing me with patience and strength to complete this thesis.

This thesis would not have been possible without the guidance, support and encouragement of numerous people around me. First of all, I would like to express my sincere gratitude to my principal supervisor Prof. Abdesselam Bouzerdoum for his enormous academic, technical and moral support. Throughout my research studies, he has been patient, zealous and very helpful to me. I owe him a profound intellectual debt because his aptitude towards research encouraged me to think and to investigate novel ideas. Special thanks go to my co-supervisor A/Prof. Son Lam Phung for his deep interest in my studies and insightful comments. He helped me generously with my writing skills and in improving the presentation of the results presented in this thesis.

I have been fortunate to receive gracious financial support for PhD studies; therefore, I would like to thank University of Wollongong for International Postgraduate Tuition Award and University Postgraduate Award. Moreover, I gratefully acknowledge the support of the staff of the School of Electrical, Computer and Telecommunications Engineering for giving me personal and professional support during my studies at the University of Wollongong. I also want to thank my current and former colleagues in the Visual and Audio Signal Processing (VASP) group. I have greatly benefited both morally and technically from casual discussions with the members of the VASP and other office mates.

I am also thankful to numerous university groups especially MAWU. Nontechnical chat with MAWU group members has always been a rich source of spiritual nourishment.

Throughout my life, I have studied from teachers who contributed a lot to my personal development. I would not be where I am without their efforts and guidance. My deepest feelings of appreciation are reserved for my parents and sisters for their persistent love and support through thick and thin. I can never pay them back for what they have contributed to each stage of my life. Finally, I want to thank my wife for her encouragement and support.

Chapter 1

Preliminaries

Chapter contents

1.1 Introduction

Much of the information about our surroundings is conveyed to us by dynamic scene events. The motion associated with any object in the scene contains very rich information, which is used in the process of decision making by humans and animals. The observation of dynamic scene events enables us to discover meaningful information, for example, speed and direction of moving objects, their properties like transparency and rigidness, three-dimensional structure of the scene, and time to collision of objects. The extraction of

meaningful information from images or image sequences is the main topic of computer vision tasks. Artificial vision systems try to describe dynamic scene events happening in the real world by analyzing and understanding visual motion in digital video sequences. Visual motion provides a rich source of information for understanding dynamic scenes. The problem of inferring the motion of objects is fundamental to computer vision as most of computer vision tasks rely on motion information. The successes of these tasks depend upon good estimates of the motion. Therefore, obtaining a good estimate of visual motion from image sequences is essential.

This chapter first provides the reader with an introduction to visual motion in digital videos and its relation to optical flow. Second, it describes research motivations by discussing major problems in optical flow estimation. Moreover, it summarizes main contributions and presents the thesis outline. Finally, publications resulted from this thesis are listed.

1.2 Visual Motion

This section presents the relation between visual motion in digital videos and optical flow. It also summarizes a few commonly used applications of motion estimation in computer vision tasks.

1.2.1 Motion Field

What do we mean by motion in digital videos? Digital videos are made up of image sequences recorded by an imaging device like a camera. These image sequences are formed when three-dimensional (3D) world points are projected onto a two-dimensional (2D) imaging surface, such as the back of a human eye or a light sensor array of a digital camera. The brightness or intensity of the formed image changes when positions of world points and objects, relative to the imaging surface, change in the 3D world. This movement of projected scene points is referred to as the *image velocity* or the *motion field*.

The motion field in digital videos is an ideal description of the actual three-dimensional

motion. In practice, only an approximation of the motion field can be determined from image sequences because it is a two-dimensional projection of the actual three-dimensional motion. In other words, this projection normally results in a data loss.

1.2.2 Optical Flow

The approximation of the motion field, in digital videos, is achieved by the mapping of visible scene points from one video frame to another. This mapping is commonly known as *optical flow*. Optical flow can be defined as a two-dimensional apparent motion field which is made up of the object motion as well as the motion of the camera. It can describe dynamic scene events by giving a pattern of the apparent motion of objects and surfaces in the scene. The motion field is generally different from optical flow as described in [7]. However, a reasonably good approximation of the motion field is provided by optical flow.

Optical flow from one video frame to another is depicted in different ways; it is sometimes shown as a quiver plot, or presented as a color image. In the quiver plot, the magnitude and the direction of any vector at a particular point gives an estimate of the motion at that point. Figure 1.1 (a) and (b) show two consecutive frames of a video sequence. The quiver plot between these two frames is given in Figure 1.1 (c). Figure 1.1 (d) presents optical flow vectors in color. The hue corresponds to the direction, whereas the saturation shows the magnitude of flow vectors at each point. The respective color scheme is shown in Figure 1.1 (e). In this thesis, we use this color scheme to illustrate optical flow vectors.

1.2.3 Applications

Optical flow has been used in a wide variety of tasks. It is used, for example, to differentiate moving objects from the stationary background in a scene. The tracking of moving objects also use optical flow; tracked objects can further be used for driving assistance systems [8], robot navigation [9], human face tracking [10] and obstacle avoidance systems [11], to name a few. Moreover, optical flow is used in abnormal activity and behavior detection [12, 13]. Motion estimation methods are extensively used in video

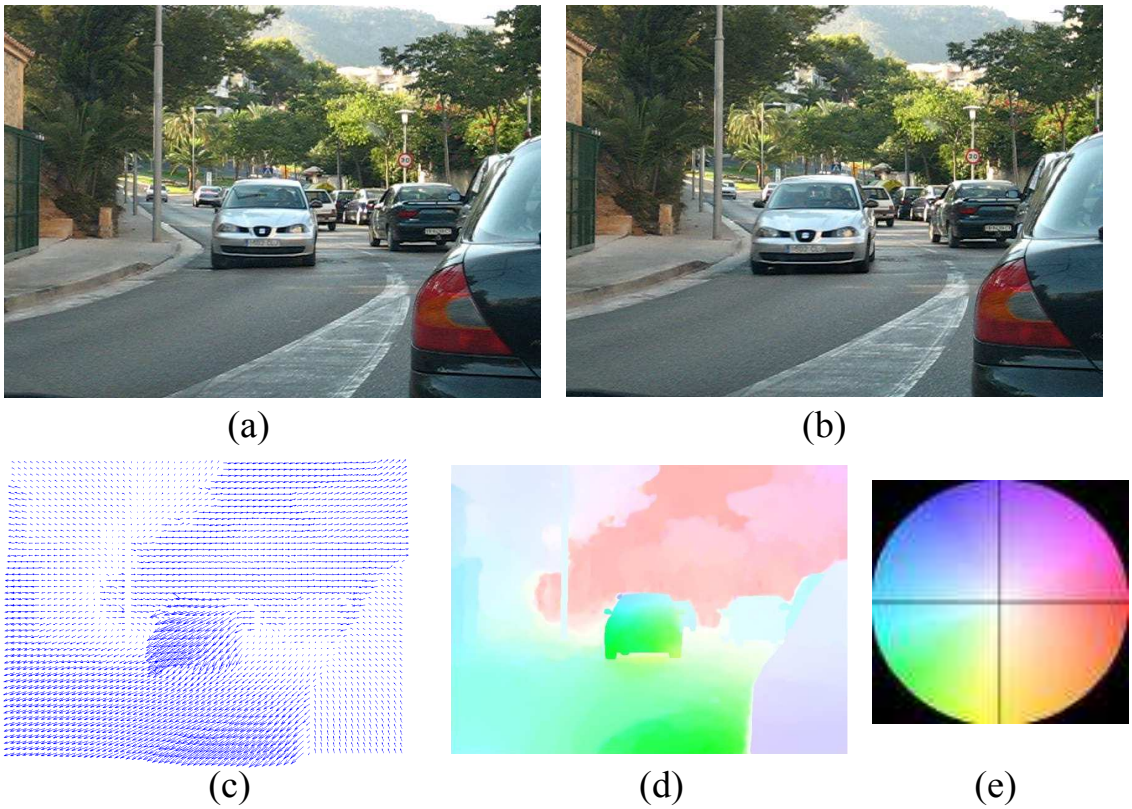


Figure 1.1: Optical flow field between (a) the reference video frame and (b) the target video frame, shown as (c) a quiver plot and (d) a color plot. The arrows in (c) and the color in (d) show the magnitude and the direction of optical flow vectors. (e) The color scheme used to show the color plot in (d).

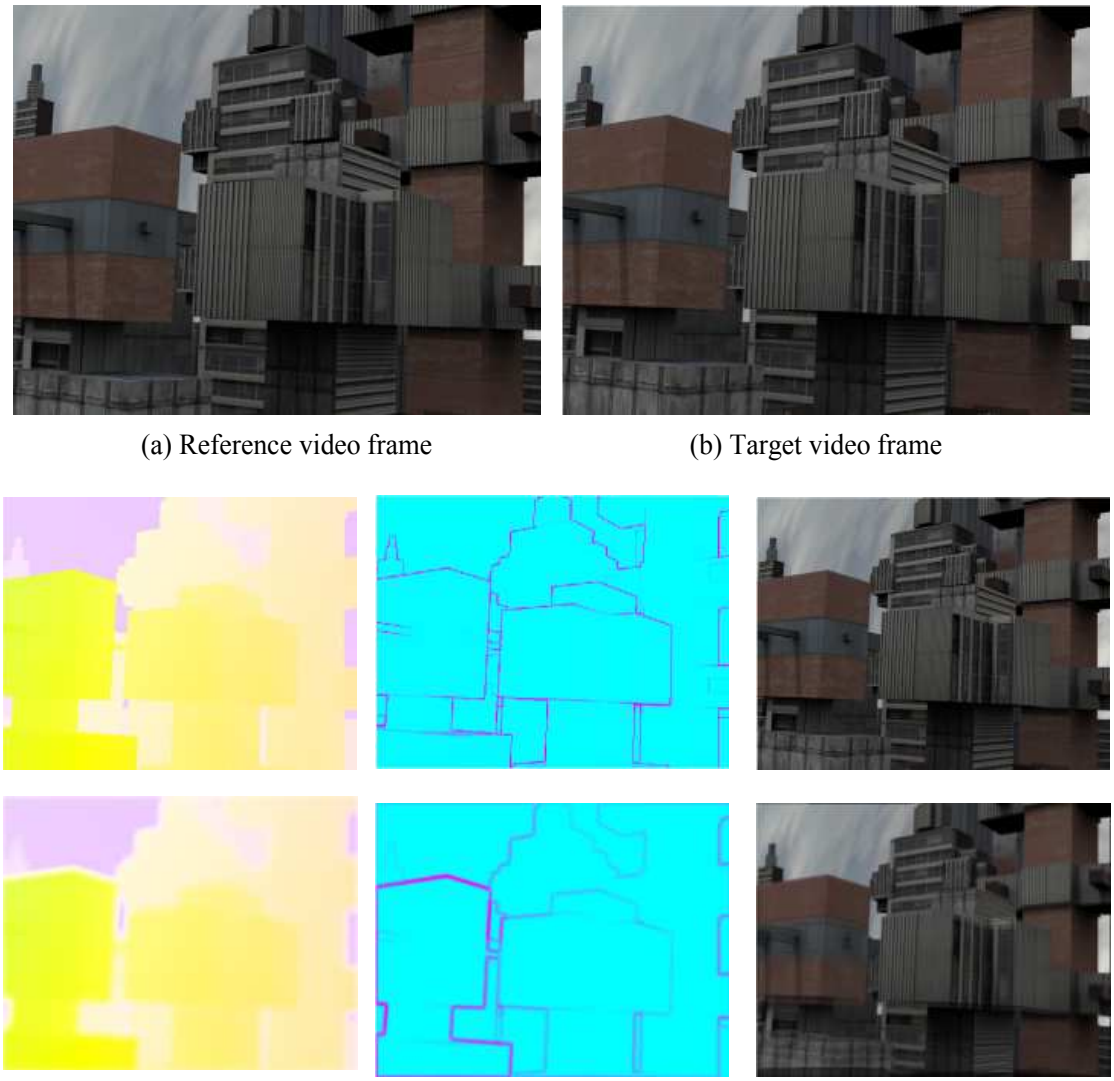
compression techniques to encode the interframe dynamic scene information as motion vectors [14]. Other examples include video stabilization [15], object segmentation [16], three-dimensional shape recovery [17], and image stitching [18]. The estimated motion of objects can be combined with machine learning techniques for human detection in a scene [19], human action recognition [20], and facial expression recognition [21], to name but a few. These techniques are usually used in human machine interfaces. Therefore, optical flow estimation is of fundamental importance to many computer vision applications.

1.3 Research Motivations

The estimation of optical flow is error-prone and remains a challenging task despite the enormous research effort spent on the topic. Several commonly occurring situations in digital videos pose trouble to most optical flow estimation methods. These methods encounter difficulties in handling motion boundaries, occluded video regions, and varying illumination of the scene. The motivation of this thesis is to understand and handle aforementioned problems. Here, we discuss these problems and their effects on performances of optical flow estimation methods.

1.3.1 Motion Boundaries

Motion boundaries are problematic for most optical flow estimation methods when special care is not taken to handle them properly. Motion boundaries occur at depth discontinuities in the scene; these convey useful information for the visual perception of the object shape and the three-dimensional structure of the scene. It is most likely that multiple objects are present at motion boundaries; thus, the motion estimation of multiple objects requires the preservation of these boundaries. Most optical flow estimation methods either blur motion boundaries or result in their delocalization when optical flow is not constrained properly. Figure 1.2 displays the effect of two different types of computed optical flows on the reconstruction of a video frame; computed optical flows and resulting motion boundaries are shown in Figure 1.2(c) and (d), respectively. In this figure, the 2nd and the 3rd rows show results by the method that preserves motion boundaries and the method that blurs motion boundaries, respectively. It can be observed that edges of the reconstructed video frame in the 2nd row are preserved, whereas same edges are blurred in the 3rd row of Figure 1.2 (e). This observation suggests that there is a strong need to handle motion discontinuities for the preservation of sharp boundaries.



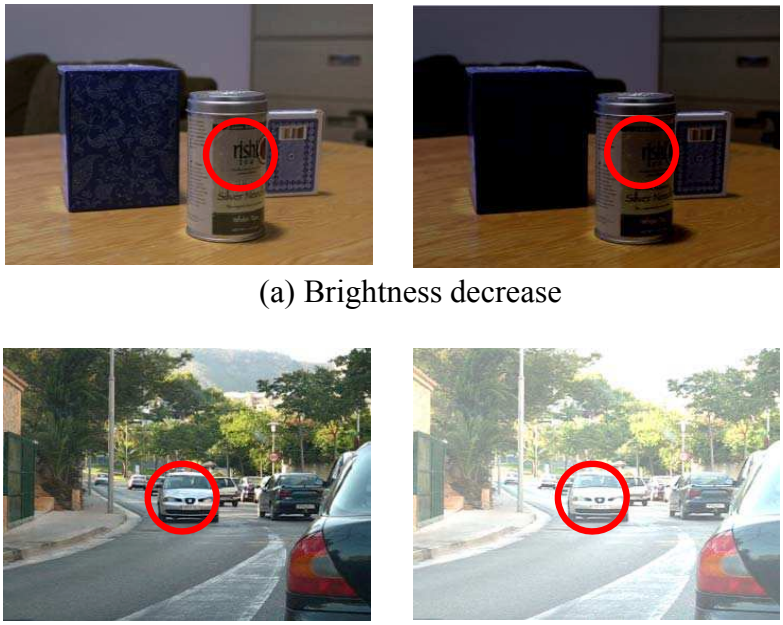
(a) Reference video frame (b) Target video frame

(c) Optical flow (d) Motion boundaries (e) Reconstructed frame

Figure 1.2: The effect of optical flow computed by two different methods on the reconstruction of digital videos. (a) The reference and (b) the target video frames. (c) Optical flow, (d) motion boundaries computed from optical flow, and (d) the reconstructed target video frame using the reference frame and optical flow. The results in (c)-(e) are shown for the method that preserves motion boundaries (2nd row) and the method that blurs motion boundaries (3rd row).

1.3.2 Illumination Change

The matching of scene points may be erroneous or even not possible during optical flow estimation when the illumination of the scene is varying with time. Consider a video



(a) Brightness decrease

(b) Brightness increase

Figure 1.3: The brightness change in digital videos: (a) the brightness decrease and (b) the brightness increase from the reference video frame (left) to the target video frame (right).

sequence obtained by imaging a scene under constant illumination. The brightness of any point in that video remains constant over time when that point does not move. However, when the illumination of the scene is changing, the brightness of a static point also changes in the video sequence. Consequently, optical flow methods may wrongly infer that the static point is moving because of wrong matching of the pixels between two consecutive video frames. Furthermore, the motion of a moving point may be estimated incorrectly when the brightness of that point changes due to varying illumination of the scene. Figure 1.3 shows the effect of varying illumination in two consecutive video frames. Scene points to be matched across video frames change their brightness due to either a decrease or an increase in the illumination of the scene. These points can not be matched by matching the brightness across the video frames. Thus, illumination variation needs to be handled properly in optical flow estimation methods for correct matching of moving patterns.

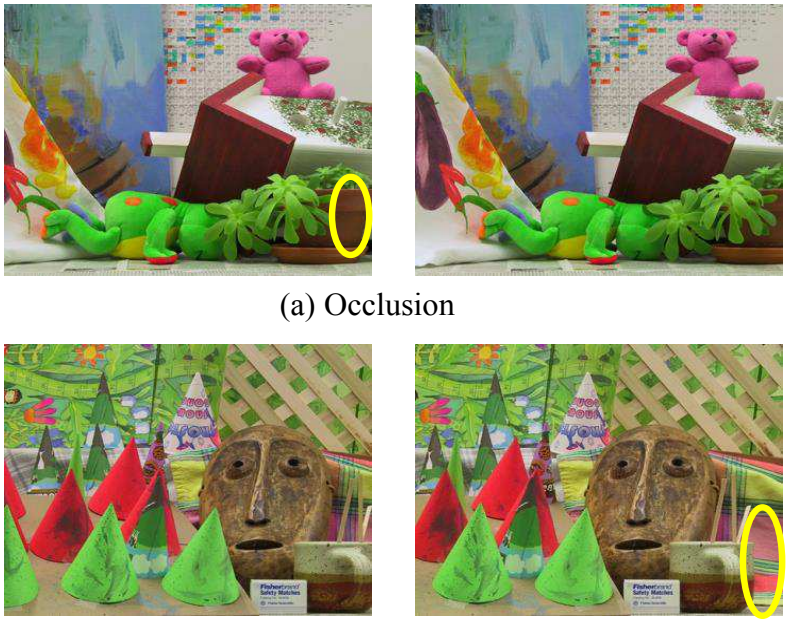


Figure 1.4: The disappearance and appearance of pixels in digital videos: the phenomena of (a) occlusion and (b) disocclusion from the reference video frame (left) to the target frame (right), as shown in the highlighted parts of frames.

1.3.3 Occlusion

The presence of occluded objects in the scene causes the appearance or disappearance of pixels in any digital video. Occluded objects can not be matched across video frames because moving patterns are undefined in occluded regions. Occlusion is considered as a bottleneck for optical flow estimation methods. These methods rely on the matching of scene points or regions across video frames. Pixels should not be matched in occluded regions because optical flow is undefined there. Figure 1.4 shows the phenomena of occlusion and disocclusion in two digital videos. The disappearance and appearance of moving patterns can be observed in highlighted parts of Figure 1.4 (a) and (b), respectively. There could be wrong matching of pixels in the reference and the target video frames when occlusions are not handled explicitly. The situation is further complicated when the illumination of the moving imagery is varying with respect to time. Optical flow estimation methods use a matching cost function; the matching of moving patterns using this

function may be wrong or may not be possible at all under varying illumination of the scene. In the absence of appropriate modeling of the brightness change of the scene, these methods can wrongly detect occlusion in visible video regions and vice versa. The wrong matching of occluded video regions significantly increases optical flow estimation error. Therefore, it is highly desirable to model these regions for their accurate and efficient detection.

1.4 Thesis Outline and Contributions

This thesis contributes to optical flow estimation by introducing numerous technical innovations to solve problems mentioned in Section 1.3. This section describes the thesis outline, and provides the chapter by chapter summary of major contributions.

Chapter 2 reviews several existing optical flow estimation methods proposed over time. Since the presented work in this thesis is based on variational methods, more focus has been put on the state-of-the-art in variational methods. The main contributions of Chapter 2 are:

- A succinct review of capabilities of existing methods to handle problems inherent to optical flow estimation. This review helps the reader understand techniques, when integrated with variational methods, which lead to the gradual improvement in the accuracy of the estimated optical flow.

- This review attempts to disclose factors, such as the choice of the objective function, weighting of different constraints, and optimization methods, which can be combined together to enhance optical flow estimation quality.

Chapter 3 briefly reviews commonly used variational regularization techniques for optical flow estimation. The emphasis is put on robust regularization techniques, particularly total variation regularization. A robust regularizer is proposed which enforces the gradient continuity of an unknown multidimensional signal in a small neighborhood.

Therefore, the proposed regularizer promotes the sparsity of partial derivatives of the computed solution. This regularizer is shown to be rotationally invariant to camera motions. Extensive experimental results are presented to show the superiority of the proposed regularizer over existing robust regularizers for the preservation of sharp boundaries. The main contributions of Chapter 3 are:

- A novel edge preserving and sparsity enhancing variational regularizer is proposed. The sparsity enhancing capability of this regularizer enables it to estimate sparse signals from a small set of available measurements.
- The proposed regularizer has been applied to problems of image denoising and image restoration from incomplete measurements. To show the effectiveness of the proposed regularizer for image denoising and image restoration, a comparison has been conducted with three existing total variation based regularizers.
- Extensive experimental results are conducted to reveal that image restoration using the proposed regularizer generally results in higher PSNR of the restored images than local TV regularizers.

Chapter 4 addresses the problem of estimating dense optical flow while preserving sharp motion boundaries. First, the regularizer proposed in Chapter 3 is integrated in a variational energy minimization framework along with various optical flow data matching terms. This regularizer exploits the sparsity of the gradient field of optical flow. Second, an efficient convex optimization solver is used to estimate a piecewise smooth flow field. Since the minimization of the proposed variational energy results in a sparse gradient field of the estimated optical flow, the problem of estimating a dense optical flow from highly incomplete measurements is also solved. The proposed method has been compared against numerous existing approaches; extensive experimental results demonstrate that the proposed method performs comparable to or better than the state-of-the-art techniques when optical flow is estimated from full or incomplete measurements, respectively. The main contributions of Chapter 4 are:

- A sound characterization of motion boundaries by the integration of the proposed regularizer of Chapter 3 into a variational energy minimization framework.
- A significant reduction in the computational complexity of variational optical flow estimation by the use of stochastic sampling. It has been shown that a reliable optical flow can be estimated from a small set of available measurements, as low as 10%, without too much sacrificing the accuracy of optical flow.
- An increased robustness of the proposed method to not only the noise but also the regularization parameter.

Chapter 5 is dedicated to occlusion detection in digital videos using optical flow, under constant and varying illumination of the scene. State-of-the-art methods in occlusion detection using optical flow are reviewed. Since these methods use the optical flow constraint (OFC) to construct data terms, the successes of these methods are based on the assumption that the OFC is not violated. The method presented in this chapter aims to accurately detect occlusion under varying illumination of the scene in a consistent variational formulation, thereby improving the quality of the estimated optical flow significantly. The proposed method treats occluded video regions as sparse because these regions form a small portion of any video frame. Furthermore, this method utilizes a generalized dynamic image model (GDIM), which can capture brightness change in its parameters. The GDIM is then used to construct a data term which explicitly models occlusion and the brightness change. A computationally less expensive version of the proposed method is also presented, which does not require the estimation of brightness change parameters of the GDIM under certain conditions. Since the proposed method exhibits robustness against illumination changes, experiments on both synthetic and real video sequences show promising results for optical flow estimation and occlusion detection. The main contributions of Chapter 5 are:

- A separate modeling of occluded regions and the brightness change. Occlusion is modeled as a sparse term because its domain is very small compared to the domain of

a video frame. The brightness change is modeled such that it is linearly parametrized in the GDIM.

- A novel estimation algorithm that jointly estimates optical flow, occlusion and brightness change parameters of the GDIM by minimizing a single variational energy.

Chapter 6 provides a summary of the problems discussed in this thesis, and proposed methods to solve these problems. It also suggests future research work after identifying some of the remaining challenges.

1.5 Publications

This section lists the articles resulted from the work presented in this thesis.

- M. W. Nawaz, A. Bouzerdoum, and S. L. Phung, "Motion estimation with adaptive regularization and neighborhood dependent constaint," in *Proceedings of the Digital Image Computing, Techniques and Analysis*, pp. 387 – 392, Dec. 1 – 3, 2010.

- M. W. Nawaz, A. Bouzerdoum, and S. L. Phung, "Optical flow estimation using sparse gradient representation," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 2681 – 2684, Sep. 11 – 14, 2011.

- M. W. Nawaz, A. Bouzerdoum, and S. L. Phung, "A novel sparse variational regularizer for optical flow estimation", (under review)

- M. W. Nawaz, A. Bouzerdoum, and S. L. Phung, "Sparse occlusion detection using optical flow under varying illumination", (under review)

Chapter 2

Literature Review

Chapter contents

2.1 Introduction

Optical flow estimation is considered as a low-level computer vision problem which is utilized in mid-level and high-level vision tasks. The computation of optical flow from

video sequences has been studied extensively because vision applications, which use motion information, greatly depend on the accuracy of the computed optical flow. This chapter reviews existing optical flow estimation methods. Several models for optical flow estimation have been proposed over time, and there are different ways to classify optical flow methods based on these models. However, this chapter abstains from any strict categorization of these methods. Methods are presented according to their model complexity and their ability to solve problems mentioned in Section 1.3.

The rest of the chapter is organized as follows. Section 2.2 presents the optical flow constraint (OFC), which is commonly used to match brightness (or intensity) patterns across video frames. Section 2.3 describes the local optical flow estimation model, and reviews local region matching and gradient methods. Section 2.4 presents the global optical flow model; it also describes global parametric, spline based and nonparametric methods. Section 2.5 discusses the role of different data and regularization terms used in variational methods. Section 2.6 presents optical flow based occlusion detection techniques. Finally, Section 2.7 concludes the chapter.

2.2 Optical Flow Constraint

Let us define a time varying video sequence $I(x, y, t)$, where I represents the intensity (or brightness) of pixels, x and y are spatial coordinates, and t is the time variable. Most optical flow estimation techniques assume that the change in the image intensity is only due to discrete displacements of pixels from one video frame to another [22]. This intensity constancy assumption between two consecutive video frames $I(x, y, t)$ and $I(x - v_x dt, y - v_y dt, t - dt)$ can be expressed as

$$I(x, y, t) - I(x - v_x dt, y - v_y dt, t - dt) = 0, \tag{2.1}$$

where dt is the temporal sampling interval, and v_x and v_y are, respectively, the horizontal and vertical components of optical flow between video frames at time t and $(t - dt)$. Equation (2.1) is nonlinear in the flow vector field; therefore, a linearized version of this

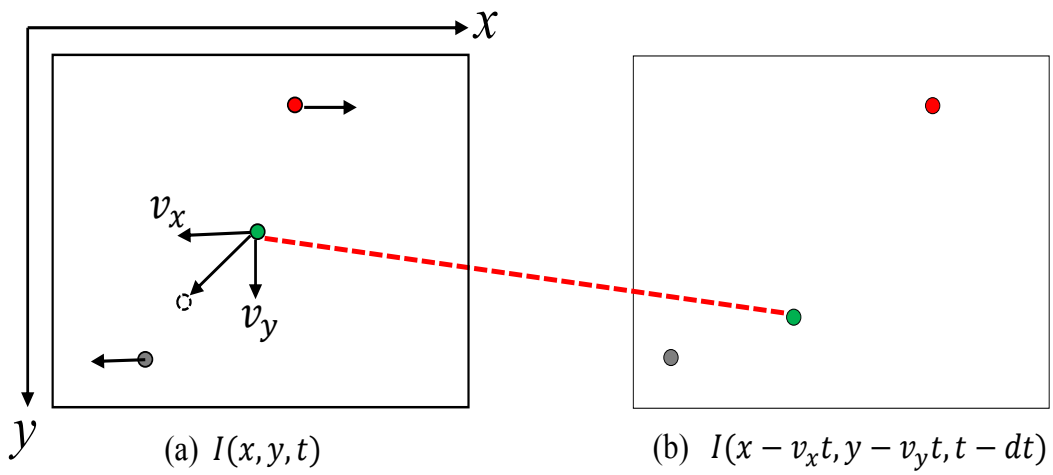


Figure 2.1: The intensity constancy assumption between (a) the reference video frame $I(x, y, t)$ and (b) the target video frame $I(x - v_x, y - v_y, t - dt)$. The pixel intensity is shown in color in both frames. Arrows in (a) are showing the direction of movement for each pixel.

assumption is usually used in a multiresolution scheme [23]. The linearized version of (2.1), also known as the *optical flow constraint* (OFC) or *brightness constancy constraint*, can be written as

$$\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} + I_t = (\nabla I)^{\mathrm{T}} \vec{v} + I_t = 0, \tag{2.2}$$

where $\vec{v} = [v_x \;\; v_y]^{\mathrm{T}}$ is a 2D velocity vector, ∇ is the gradient operator, $\nabla I = [I_x \;\; I_y]^{\mathrm{T}}$ and I_t are, respectively, spatial and temporal intensity derivatives. Note that we have omitted the spatiotemporal dependency of terms in (2.2) for the sake of brevity. Intensities of moving pixels from reference video frame $I(x, y, t)$ to target frame $I(x - v_x, y - v_y, t - dt)$ are preserved as shown in Figure 2.1. Moving pixels follow Equation (2.2) when their intensities do not change over time. The OFC is used for the pixel matching across video frames by most optical flow estimation methods. A variant of the OFC or advanced matching terms can be employed when the brightness constancy assumption is violated.

The OFC is a pointwise constraint. It can only calculate the local motion component that is normal to an image structure or parallel to the intensity gradient. The motion of an edge or a contour is inherently ambiguous when observed locally. This local motion is consistent with numerous possible directions. The perceived direction of local motion of

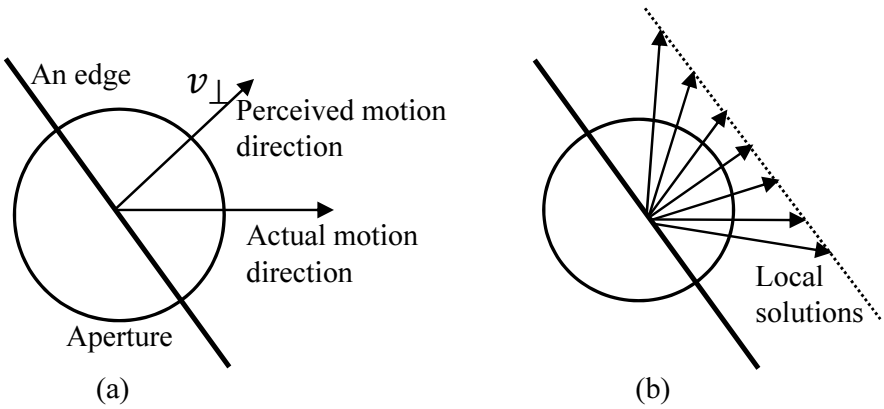


Figure 2.2: The aperture problem: (a) an edge translating to the right as seen through a circular aperture; (b) local solutions to the aperture problem, which are consistent with numerous possible motion directions.

an object is perpendicular to the orientation of that object for any motion direction. This problem is widely known as the *aperture problem*. The direction of motion of an object is indistinguishable when seen through an aperture. The pointwise OFC at each pixel acts like an aperture; thus, it computes optical flow at each pixel independently of other pixels. A similar observation has also been reported in motion sensors present in the visual cortex of mammals [24]. Since these motion sensors have finite receptive fields, they are capable of detecting only local motion of objects. Figure 2.2 shows an edge, translating from left to right, viewed through a circular aperture. This left to right motion produces the same spatiotemporal pattern as a set of lines moving from bottom to top. Hence, this phenomenon creates the ambiguity in determining the actual motion. To cope with the aperture problem, extra information needs to be incorporated. This could either be local integration of the image information or the inclusion of a global regularization term along with the OFC.

2.3 Local Optical Flow Model

This section presents the local model to avoid aperture problem in optical flow estimation methods. The local optical flow model uses a few local measurements of image intensity in

a neighborhood. There are two major types of methods based on the local model: region matching methods and gradient methods. These two types of methods, in general, are able to estimate a single motion in a local neighborhood. Some variants of both methods have been proposed which can estimate multiple motions as well. Here, both types of methods are discussed for single and multiple motion estimation.

2.3.1 Region Matching

Region or patch-based matching methods are simple and easy to implement. These methods assume that there is only a single motion in the patch. This type of motion arises when a plane moves in front of a camera [25]. Region matching methods are widely used in video compression techniques because the goal is to achieve the highest possible compression ratio. In these methods, a similarity measure is chosen to calculate the resemblance between image patches. The sum of the squared differences (SSD) between image patches can be minimized to achieve the maximum similarity:

$$E(\vec{v}) = \min_{\vec{v}} \sum_{\Re} \Big[I(x, y, t) - I(x - v_x dt, y - v_y dt, t - dt) \Big]^2, \tag{2.3}$$

where $\Re$ represents a local neighborhood, and $E(\vec{v})$ denotes the energy between image patches which is sought to be minimized. After choosing an image region $\Re$, a search for the pixels of the template image $I(x, y, t)$ is conducted in the target image $I(x - v_x, y - v_y, t - dt)$ in such a way that the similarity measure is maximized. The estimated motion vector, which maximizes the similarity measure, is then used to replace a matched patch across video frames.

The cross-correlation serves as a substitute for image intensity difference to calculate the similarity between image patches [26, 27]. A method using the cross-correlation seeks to maximize the product of patches across video frames as

$$E(\vec{v}) = \max_{\vec{v}} \sum_{\Re} \Big[I(x, y, t) I(x - v_x dt, y - v_y dt, t - dt) \Big]. \tag{2.4}$$

Methods using either the intensity difference or the cross-correlation assume that the intensity of consecutive video frames to be matched does not change over time. However, the

intensity constancy assumption may not hold across video frames whenever the exposure is varied or the illumination of the light source is changed during the video acquisition.

The normalization of the intensity constraint provides robustness against varying illumination [28, 29, 30]. For color images, color channels can also be normalized to mitigate the effect of the brightness variation [31, 28]. Similarly, the normalized cross-correlation (NCC) is substituted for the cross-correlation in some methods as in [32, 33]. In an image patch, let $\overline{I}(x, y, t)$ and $\overline{I}(x - v_x dt, y - v_y dt, t - dt)$ be mean intensities for $I(x, y, t)$ and $I(x - v_x dt, y - v_y dt, t - dt)$, respectively. Furthermore, let us define the mean subtracted pixel intensities in $I(x, y, t)$ and $I(x - v_x dt, y - v_y dt, t - dt)$ as

$$\widehat{I}(x, y, t) = I(x, y, t) - \overline{I}(x, y, t), \tag{2.5}$$

and

$$\widehat{I}(x - v_x dt, y - v_y dt, t - dt) = I(x - v_x dt, y - v_y dt, t - dt) - \overline{I}(x - v_x dt, y - v_y dt, t - dt). \tag{2.6}$$

The NCC is now given as

$$E(\vec{v}) = \max_{\vec{v}} \frac{\sum_{\Re} \left[\widehat{I}(x, y, t)\right] \left[\widehat{I}(x - v_x dt, y - v_y dt, t - dt)\right]^2}{\sqrt{\sum_{\Re}[\widehat{I}(x, y, t)]} \sqrt{\sum_{\Re}[\widehat{I}(x - v_x dt, y - v_y dt, t - dt)]}}. \tag{2.7}$$

Note that intensities of image patches in (2.7) are normalized by subtracting mean intensity values. This normalization makes the NCC robust against variations in intensity but it increases computations.

The cross-correlation or the NCC are more expensive to compute than the intensity difference. An efficient way to compute the cross-correlation or the NCC is to take the Fourier transform of both video frames to be matched. A correlation function appears as a multiplication in Fourier domain which can significantly accelerate computations [34]. For an n pixel video frame, the complexity of operations is thus reduced from n^2 to $n\log(n)$ using the fast Fourier transform (FFT). The complexity of the cross-correlation can also be decreased in the spatial domain by using box filtering techniques [35], which are invariant to the size of correlation windows. In the spatial domain, partial elimination

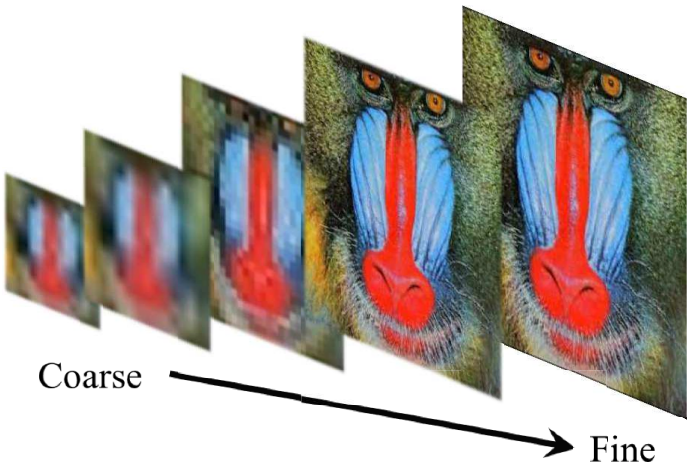


Figure 2.3: A five level coarse-to-fine image pyramid representing a single image at different resolutions.

technique has been employed for early termination of the template matching calculation [36]. This technique stops matching computations at a particular search location when the match is less reliable than an already computed best matching candidate location [37]. The computational advantage of partial elimination techniques over frequency domain techniques may not be significant for large template sizes. However, when the template size reduces, spatial domain elimination techniques perform better than frequency domain techniques.

Performances of region based matching techniques deteriorate for large inter-frame motion of objects. The estimation of large motions requires large matching windows which results in a computationally intensive search within these windows. Selecting a large neighborhood is not a feasible solution. There is also a potential risk of mismatch in the big search space. In order to reduce the burden of calculations, hierarchical approaches are used. Since large motions appear small at low resolution, an image pyramid is constructed. Optical flow is then estimated at each level of the pyramid. The estimated optical flow at a low resolution level of the pyramid serves as an initial guess for images at next fine level. Figure 2.3 shows five levels of such a pyramid which contains the same image at different resolutions.

In [23], Anandan employs a Laplacian pyramid to describe a coarse-to-fine matching strategy for a dense optical flow estimation. The image intensity at coarser levels of the pyramid is used to obtain initial motion estimates, which are then refined by using small scale intensity information at higher resolution levels of the pyramid. This method first uses a quadratic approximation of the SSD surface for the patch matching. A confidence measure is then derived by computing the curvature at the minimum of the SSD surface. Estimated optical flow vectors are then propagated into less textured video regions according to this confidence measure. Another hierarchical method, presented by Singh [38], uses two stages for the matching of video frames. This method requires three video frames in contrast to two video frames used by Anandan. The SSD is computed in the first stage on highpass filtered images. The SSD surface is then converted into a probability distribution. Subpixel motion estimates are obtained by computing the weighted mean of this distribution. In the second stage, a smoothing term is used which is weighted according to covariance matrices associated with velocity estimates of the first stage. Singh's method generates a dense optical flow estimate by propagating the estimated motion vectors into low-textured video regions.

2.3.2 Gradient Methods

The OFC expresses optical flow in terms of the spatiotemporal gradient. At each pixel, the OFC constrains optical flow vector to a line in the (v_x, v_y) plane. Every pixel in a small neighborhood has its OFC line. Multiple OFC lines associated with pixels in a small neighborhood can be integrated to over-constrain optical flow. The resulting over-constrained system of equations resolves the aperture problem. These equations have been solved by minimizing the error in a weighted least squares approach by Lucas and Kanade [39]:

$$E(\vec{v}) = \min_{\vec{v}} \sum_{\Re} W(x,y)\Big(\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} v_x & v_y \end{bmatrix}^{\mathrm{T}} + I_t \Big) = \min_{\vec{v}} \sum_{\Re} W(x,y)\Big((\nabla I)^{\mathrm{T}} \vec{v} + I_t \Big), \quad (2.8)$$

where $W(x,y)$ is a weighting function which usually decreases with the distance from the center of the neighborhood. The least squares solution can be written in the matrix form

as

$$\begin{bmatrix} \sum_{\Re} W(x,y) I_x^2 & \sum_{\Re} W(x,y) I_x I_y \\ \sum_{\Re} W(x,y) I_x I_y & \sum_{\Re} W(x,y) I_y^2 \end{bmatrix} \vec{v} = - \begin{bmatrix} \sum_{\Re} W(x,y) I_x I_t \\ \sum_{\Re} W(x,y) I_y I_t \end{bmatrix}. \tag{2.9}$$

This local gradient method produces one flow vector in each image patch by assuming that optical flow is constant. Since the method of Lucas and Kanade is based on the linearized version of the brightness constancy assumption, it is limited to small displacements only. Nagel [40] and Uras et al. [41] replaced the first-order Taylor series approximation of the brightness constancy with the second-order approximation to handle large motions. The second and higher-order derivatives are noisy, and it is likely that the noise in their estimated values would overshadow their benefit. Therefore, these techniques are rarely used in practice as pointed out by Baker and Matthews [42]. A coarse-to-fine pyramidal approach is employed for large displacements in [43]. This method uses down-sampled image sequences to find an initial estimate of optical flow. This estimate is utilized as initial guess for final optical flow estimation.

The OFC is inherently noisy and error-prone because it is based on the brightness constancy assumption which may be violated. The least squares approach is optimal when the brightness change is zero-mean Gaussian, and independent and identically distributed (i.i.d.). In case of the brightness constancy violation, the i.i.d. assumption does not hold. Therefore, instead of a least squares penalty, a robust measure is needed. Techniques from robust statistics are sometimes applied to acquire the robustness against noisy OFC measurements. The least median of squares regression performs significantly better than the ordinary least squares when the OFC is violated [44]. Total least squares has also been used to obtain a robust flow estimate after pre-filtering video sequences with multiscale filters of different orientations and frequency responses [45]. Multiple constraints are thus obtained in the neighborhood which are solved using the total least squares.

Gradient methods are computationally less expensive than most of local optical flow estimation methods. In practical situations, a fast and real time algorithm is desirable. Therefore, gradient methods are widely used in the tracking of the moving objects [46, 47, 48, 49]. Moreover, if the source and target video frames appear to be similar, the gradient

of the second warped frame can be replaced by the first frame in (2.8). Thus, a speed up of gradient methods can be achieved by pre-computing gradients of the source frame instead of the warped target frame. This results in a substantial decrease in computations. This technique is first developed by Hager and Belhumeur [47], and then refined by Baker and Mathews [42]. The inverse of the estimated incremental motion is added to the previously estimated motion; therefore, it is named as *inverse compositional algorithm*. A further speed up of about 15 percent is achieved by utilizing mutual information in place of the intensity difference in the algorithm of Dowson and Bowden [50].

Probabilistic techniques have been used with gradient methods to handle uncertainty in optical flow estimation. The noise in image acquisition, the aperture problem, the brightness change and the presence of multiple motions in a small neighborhood are a few of many factors causing this uncertainty. In [51], Simoncelli models these uncertainties by utilizing the image gradient and the probability distribution of the motion. A Gaussian distribution is assumed over these gradients. The motion is then taken as the maximum likelihood estimate (MLE), which is the mean of this distribution. The computation of the MLE of a Gaussian distribution is the same as the linear least squares estimation. This model works well for normally distributed motion field errors; however, sophisticated models are needed when these errors have different distributions. In [52], Govindu uses a Fuzzy bow-tie distribution, described in [53], on brightness gradients. This representation is integrated with the method of Lucas and Kanade as well as with the affine optical flow model. The resulting method is then used to estimate a normalized version of multiframe optical flow in a spatiotemporal volume.

2.3.3 Multiple Motion Estimation

The underlying assumption behind local methods is that the motion is locally constant. Therefore, these methods produce a single optical flow vector in the neighborhood region $\Re$. The constancy of optical flow in a spatial neighborhood is a restrictive condition for many common situations in videos. Motion discontinuities and occlusion boundaries are

particularly problematic for methods utilizing this assumption. Multiple motions are expected to be observed at motion discontinuities and occlusion boundaries. One way to handle multiple motions is to reduce the size of the neighborhood in local methods. In [54], Okutomi and Kanade use a probabilistic approach to adapt the size of the matching window in a region-based matching approach by analyzing the uncertainty of the computed motion. Since adapting the window size merely decreases the size of the rectangular window in the vicinity of multiple motions, this approach does not really address the underlying problem.

Some local methods are based on the idea of first segmenting or over-segmenting images by using the intensity information before actual motion estimation [55, 56]. However, in these methods, the quality of the final estimated motion depends upon initial segmentation results. When the segmentation results are erroneous, the optical flow estimates will be wrong in subsequent stages after segmentation. In [57], hard thresholding is replaced by soft thresholding for over-segmentation of regions to minimize the effect of wrong segmentation. This method is based on the framework of Malik et al. [58], which groups sampled image points by using an image based affinity metric. Thus, in this method, the use of soft thresholding reduces the effect of initial segmentation on the final estimated motion.

Robust statistics can also be used for the estimation of multiple motions in local methods. Suppose, for example, that there are two motions in a small neighborhood. The pooling of constraints in the neighborhood will result in two different types of constraints. The first type of constraint is consistent with one of the motion, while the second type is associated with the other motion. The constraint of the first type of motion appears as an outlier for the second type of constraint. To detect these outliers for multiple motion estimation, a statistical framework is used in a gradient based method by Black and Anandan [59]. Their method replaces the ℓ_2 penalty function with robust Geman-McClure and Lorentzian penalty functions to detect multiple motions in a local window.

2.4 Global Optical Flow Model

The global optical flow model incorporates measurements over the entire image frame by defining a global energy functional. Methods based on the global model solve for optical flow by minimizing this energy functional. These methods have gained more popularity for their ability to produce higher quality results compared to their local counterparts. A variety of global techniques has been proposed over time. Here, global methods are discussed according to their complexity and their capabilities to handle commonly encountered optical flow estimation problems mentioned in Section 1.3. Global methods, in a very loose sense, can be described as being parametric, semi-parametric (spline based) or nonparametric (variational). The following discussion highlights the advantages and limitations of global methods.

2.4.1 Parametric Methods

The assumption of constant motion in local methods is attractive due to its simplicity and moderate computational cost. This restrictive assumption is not suitable for many tasks which require more sophisticated models. Consider, for example, the video sequence generated when a moving camera is panned through the scene. The motion of image points can be described by an affine transformation in this case. Therefore, a parametrized motion vector seems to be more appropriate, instead of a constant translational motion vector. This parametrized motion vector describes the spatially varying motion field. The corresponding parameters of the motion field are then sought to be estimated rather than estimating the motion vectors themselves. In [60], Bergen et al. use parametric motion model for affine, planar surface and rigid body optical flows. The affine motion can be described by a six element vector $\mathbf{a} = [a_1\ a_2\ a_3\ a_4\ a_5\ a_6]$ as follows:

$$v_x(x, y) = a_1 + a_2 x + a_3 y, \tag{2.10}$$

$$v_y(x, y) = a_4 + a_5 x + a_6 y, \tag{2.11}$$

where x and y are spatial coordinates. The parameter vector $\mathbf{a}$ is then directly estimated in place of the motion field. The final motion estimate is obtained by using (2.10) and (2.11). The parametric motion model is usually embedded into a global motion estimation method; however, it is important to note that this model can also be applied to a small neighborhood in a local method.

The affine motion model can over-constrain the system of equations when applied to the entire image domain containing multiple moving objects. Therefore, a quadratic parametric motion model is preferable to represent the motion [60, 61, 62]. The quadratic motion model accounts for more complex motions than what an affine model can represent, such as the planar surface flow. It can be modeled by an eight element parameter vector $\mathbf{a} = [a_1\ a_2\ a_3\ a_4\ a_5\ a_6\ a_7\ a_8]$ as follows:

$$v_x(x, y) = a_1 + a_2 x + a_3 y + a_7 x^2 + a_8 xy, \tag{2.12}$$

$$v_y(x, y) = a_4 + a_5 x + a_6 y + a_7 xy + a_8 y^2. \tag{2.13}$$

The quadratic motion model is valid for planar surfaces and rigid objects. Robust estimation techniques have also been applied to parametric motion estimation methods [61, 63, 64]. The quadratic model along with robust error norms yields accurate flow fields when a camera moves freely in a static rigid world where there are no large 3D motions.

The parametric model is also applied to 3D motion estimation when a suitable camera representation is used. In [65], Irani et al. use plane plus parallax representation to recover the motion and the three-dimensional structure relative to the planar surface in the scene. A fairly accurate motion estimate is obtained provided there is a sufficient three-dimensional structure present in the scene. The problem becomes ill-conditioned for primarily planar or distant scenes because the basic quadratic motion model is applicable to the planar surface. In [66], a method called Q-Warping is proposed to deal with the quadric reference surface. Q-warping generates a residual optical flow proportional to the three-dimensional deviation of the surface from a virtual quadric surface. It has been successfully employed to find the warping between two views of an object.

The computational cost in parametric models is higher than in models with pure translational motion due to increased number of parameters. For a motion constrained by p parameters in an n pixel video frame, the computational complexity is $O(p^2n)$. The computational cost of these methods can be significantly reduced by dividing the video frame into smaller patches, and performing pixel level operations within small sub-blocks [67].

2.4.2 Spline based Methods

Parametric motion models represent image motion with a small number of parameters. These methods are applicable to the planar surface flow; therefore, they can be used in applications such as video stabilization where the camera motion has to be removed. A large number of parameters is required for relatively complex motion patterns. These motion patterns of image points can be modeled in terms of smooth polynomial functions, widely known as *splines*. Splines are piecewise smooth polynomial functions. Methods which use splines constrain optical flow less than parametric methods. Thus, spline based methods lie in between parametric and nonparametric optical flow methods; the latter are discussed in the next section. Splines represent optical flow in terms of spline basis functions which are controlled by control nodes [68]:

$$\vec{v}_i = \sum_j = \hat{v}_j S_j(x_i, y_i), \tag{2.14}$$

where $S_j(x_i, y_i)$ are spline basis functions, $\hat{v}_j$ are control nodes and $\vec{v}_i = [v_{xi}\ v_{yi}]^{\mathrm{T}}$. Figure 2.4 (a) shows an example of a spline motion field. The spline representation can be used with any error metric, for example, the SSD error metric given in (2.3). The resulting system of equations is similar to parametric motion equations given in Section 2.4.1.

Splines are simple to construct but they can approximate fairly complex motions due to their curve fitting and smoothness capabilities. Spline based motion estimation methods impose smoothness on the motion field; therefore, these methods remove the need for any additional regularization such as a smoothness constraint in many situations. However, in the presence of large low-textured video regions, it becomes essential to impose an extra

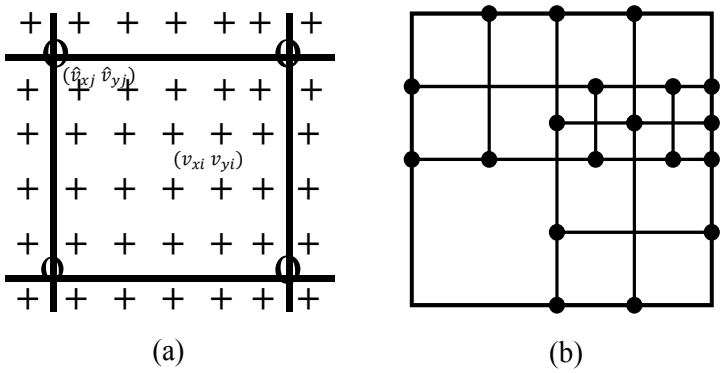


Figure 2.4: The spline motion field: (a) motion vectors $\vec{v}_i = (v_{xi}\ v_{yi})$ are shown as pluses which are controlled by control nodes $\hat{v}_j = (\hat{v}_{xj}\ \hat{v}_{yj})$ shown as circles [1]. (b) A quadtree spline grid which is used to control the size of cells in the control grid.

regularization to avoid the aperture problem [69]. Spline based methods allow for local deformations, and lead to a compact motion representation. However, particular care has to be made to handle motion discontinuities. Since a small number of control vertices is usually used in these methods, the estimation at flow discontinuities can be of poor quality, unless a large number of vertices is used. One way to mitigate the effect of this problem is to use a quadtree spline representation [70]. Such a representation allows the use of variable cell sizes in the control grid for different video regions. Applying extra number of cells near motion boundaries consequently improves the quality of results in these regions. Figure 2.4 (b) shows an example of a quadtree spline control grid with different cell sizes. Since they allow local deformations, spline based methods are attractive for a number of applications especially nonrigid medical image registration [71, 72, 73, 74]. The interested reader is referred to [75] for a performance evaluation of several optimization algorithms for medical image registration.

2.4.3 Variational Methods

The parametric and spline based methods are practicable for image registration due to their simplicity. However, these methods can not handle very complex motions. Their performances deteriorate at depth and motion discontinuities because they over-constrain

the motion field in these regions. The most general class of global optical flow methods is based on nonparametric variational model. The methods based on the variational model compute an independent and a dense estimate of the motion field at each pixel. The basic variational model, originally proposed by Horn and Schunck [22], seeks to minimize an energy $E(\vec{v})$ of the form:

$$E(\vec{v}) = \min_{\vec{v}} E_{\text{data}}(\vec{v}) + \lambda E_{\text{reg}}(\vec{v}), \tag{2.15}$$

where $E_{\text{data}}(\vec{v})$ is the data fidelity term, $E_{\text{reg}}(\vec{v})$ is the regularization term, and λ is a regularization parameter. A matching cost function is used to construct a data term which can be based on the OFC or its variants. A smoothness constraint is used as a regularization term, which resolves the aperture problem.

Due to its elegant mathematical formulation, the variational optical flow model, can embed application specific knowledge, such as the expected noise level in video sequences [76, 77], and the brightness change [78, 4, 30]. Furthermore, robust norms [79, 80] and discontinuity preserving regularizers [2, 81, 82, 83] can easily be employed. In variational methods, the energy is usually minimized in the spatial domain; thus, these methods should not be confused with the energy based optical flow methods which extract the motion energy in the frequency domain [84, 85, 86, 87]. Several variants of the basic variational optical flow model of Horn and Schunck have been proposed. Providing a detailed review of these variants is beyond the scope of this thesis. The interested reader is referred to [1, 88] for an in-depth study of variational methods. In the following section, we give an overview of data and regularization terms which have been used to solve problems mentioned in Section 1.3.

2.5 Data and Regularization Terms in Variational Methods

Recent advancements in the design of data and regularization terms have led to significant improvements in the performance of optical flow estimation methods. Here, the techniques

and the practices which result in improved estimation quality are discussed with respect to the data and regularization terms.

2.5.1 Data Term

Variational optical flow methods usually construct the data term by using the optical flow constraint (OFC). The first variational method proposed by Horn and Schunck [22] uses a quadratic data term that can be expressed as

$$E_{\text{data}}(\vec{v}) = \sum_{\mathfrak{D}} (I_x v_x + I_y v_y + I_t)^2, \tag{2.16}$$

where $\mathfrak{D}$ is the discrete image domain. The linearized OFC is valid only for small displacements. Therefore, similar to the other optical flow methods, an image pyramid is constructed to handle large displacements [89, 90]. The OFC is then used in a coarse-to-fine warping scheme to estimate large displacements.

The OFC is inherently noisy and a model that incorporates the quadratic penalty of the OFC can not cope with the deviations from model assumptions. Techniques from robust statistics are popular for handling noisy measurements. Thus, the quadratic penalization of the OFC is usually replaced by robust error functions. In [59], Black and Anandan use robust Geman-McClure and Lorentzian penalty functions which are given, respectively, as

$$\rho(x, \epsilon) = \frac{x^2}{x^2 + \epsilon}, \tag{2.17}$$

$$\rho(x, \epsilon) = \log[1 + \frac{1}{2}(\frac{x}{\epsilon})^2], \tag{2.18}$$

where the parameter ϵ is used to control the amount of penalty on values of x. The robust version of the OFC is then given as

$$E_{\text{data}}(\vec{v}) = \sum_{\mathfrak{D}} \rho(I_x v_x + I_y v_y + I_t, \epsilon). \tag{2.19}$$

The use of aforementioned robust penalty functions makes the data term nonconvex. The graduated nonconvexity can be used for the convex approximation of resulting energy functionals [79]. A better approach is to use a convex ℓ_1 norm with the data term as

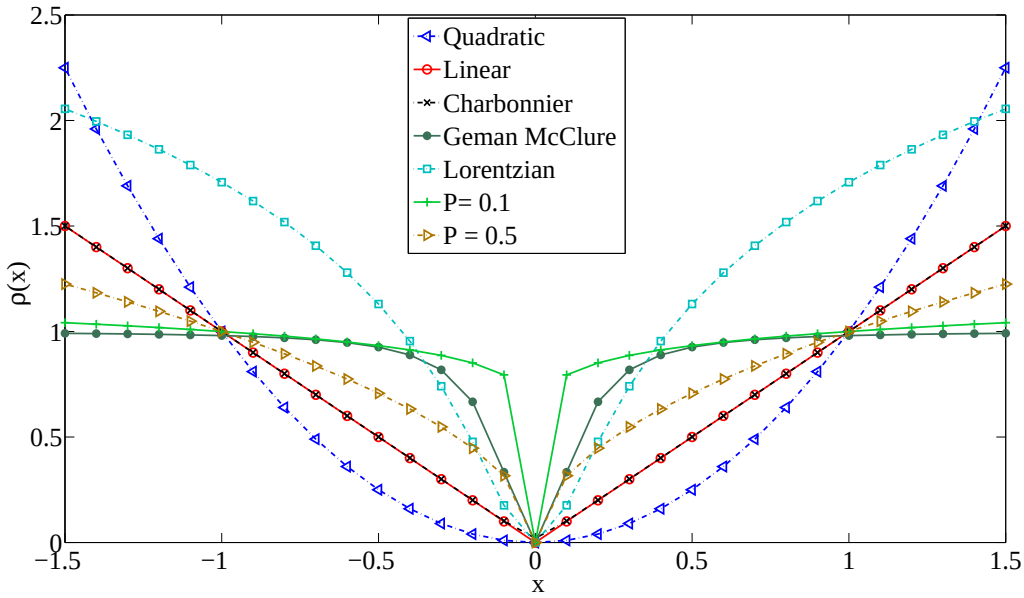


Figure 2.5: Commonly used penalty functions for data and regularization terms of variational optical flow methods. Semi-ℓ_p norms of an arbitrary signal x are also shown for $p = 0.1$ and $p = 0.5$.

employed in many recent optical flow methods [91, 92, 4, 30, 82, 93]. Bruhn et al. [76] and subsequently Sun et al. [4] use robust Charbonnier penalty, which is a differentiable approximation of the linear penalty. It is given as

$$\rho(x, \epsilon) = \sqrt{x^2 + \epsilon^2}. \tag{2.20}$$

Figure 2.5 shows commonly used penalty functions in variational optical flow methods. It can be observed that aforementioned penalty functions heavily penalize small magnitude of x and vice versa. Therefore, these functions promote the sparsity of the computed solution because it is well known that the ℓ_p norm minimization generates sparse solutions when $p \leq 1$.

Since the brightness constancy assumption is frequently violated in many real video sequences, the OFC should not be used to map points across consecutive video frames when various radiometric factors cause the change in the scene radiance and irradiance. One alternative to the OFC is to impose the constancy of the image intensity gradient

[94, 91, 95]. A data term using this constancy assumption can be expressed as

$$E_{\text{data}}(\vec{v}) = \sum_{\mathfrak{D}} \Big(\nabla I(x,y,t) - \nabla I(x - v_x dt, y - v_y dt, t - dt)\Big)^2. \tag{2.21}$$

The constancy of higher-order image derivatives can also be used, for example, Laplacian and Hessian constancy [78]. However, it is worth mentioning that these derivatives are hard to estimate. The noise in estimated values of higher-order derivatives can deteriorate results instead of introducing any improvement. Moreover, it has been recently shown in [82] that in some areas of video sequences with the brightness constancy violation, selection of one appropriate constancy constraint performs better rather than combining multiple constancy constraints together to form the data term. Aforementioned constancy assumptions are mostly used on gray level images. The natural way to extend these constraints on color images is to apply them separately on each channel [96], or use a color space that has a distinct level of photometric invariance greater than RGB, for example, HSV color space [30].

Physics based brightness change models are sometimes used to replace the OFC for complex brightness change patterns resulting from physical phenomena [97]. These models can be approximated by a generalized dynamic image model (GDIM) for two frame motion estimation as proposed by Negahdaripour [98]. The GDIM can be given as

$$I(x,y,t) - I(x - v_x dt, y - v_y dt, t - dt) = d(x,y,t)I(x,y,t) + c(x,y,t), \tag{2.22}$$

where the multiplier field $d(x,y,t)$ permits the change in the image contrast while the offset field $c(x,y,t)$ allows for the mean brightness change. Thus, a data term using the GDIM captures the brightness change as an affine function of the image intensity, and it exhibits robustness against this change.

2.5.2 Regularization Term

A simple form of flow field regularization is the quadratic regularization used by Horn and Schunck [22], which can be expressed as

$$E_{\text{reg}}(\vec{v}) = \sum_{\mathfrak{D}} (|\nabla v_x|^2 + |\nabla v_y|^2). \tag{2.23}$$

This regularization with the use of the quadratic penalty applies the same amount of weighting of the smoothness constraint throughout the image domain $\mathfrak{D}$. Such type of regularization blurs motion boundaries and discontinuities. Different methods have been proposed to handle the blur problem associated with the quadratic penalization of flow field variations. Some methods use the structure of the underlying image to vary the weighting of the smoothness term. In [99], Nagel introduced the concept of the oriented smoothness which performs anisotropic diffusion on the gradient flow. The oriented smoothness applies the smoothness along motion boundaries but inhibits it across them. Mathematically, this can be given as

$$E_{\text{reg}}(\vec{v}) = \sum_{\mathfrak{D}} (\nabla \vec{v}^{\mathrm{T}} W(\nabla I) \nabla \vec{v}), \tag{2.24}$$

where $\nabla \vec{v} = (\nabla v_x \ \nabla v_y)^{\mathrm{T}}$ and $W(\nabla I)$ is a weighting function. It can be given as

$$W(\nabla I) = \begin{pmatrix} I_y \\ -I_x \end{pmatrix} \begin{pmatrix} I_y \\ -I_x \end{pmatrix}^{\mathrm{T}} + \alpha \begin{pmatrix} I_{yy} & -I_{xy} \\ -I_{xy} & I_{xx} \end{pmatrix} \begin{pmatrix} I_{yy} & -I_{xy} \\ -I_{xy} & I_{xx} \end{pmatrix}^{\mathrm{T}}, \tag{2.25}$$

where the weighting of terms is controlled by a positive constant α. In [100], Alvarez et al. used an image driven regularization given by

$$E_{\text{reg}}(\vec{v}) = \sum_{\mathfrak{D}} W(\nabla I)(|\nabla v_x|^2 + |\nabla v_y|^2), \tag{2.26}$$

where the weighting function $W(\nabla I)$ is a positive decreasing function of the image intensity gradient. Weighted regularization generally performs better than regularization without weighting. Therefore, modern state-of-the-art methods use the regularization weighted by the image gradient [82, 2, 5, 101]. Since image driven regularization use intensity gradient information to control the amount of the regularization, it avoids the undesirable smoothing of only those motion discontinuities which coincide with image discontinuities. When motion discontinuities are different from image discontinuities, these methods produce artefacts in highly textured regions.

To cope with the situation where image and flow boundaries do not coincide, some methods use the flow driven regularization to respect motion discontinuities [95, 102]. A

flow driven regularization is of the form

$$E_{\text{reg}}(\vec{v}) = \sum_{\mathfrak{D}} \varphi(|\nabla v_x|^2 + |\nabla v_y|^2), \tag{2.27}$$

where $\varphi(x^2)$ is a convex increasing function of x:

$$\varphi(x^2) = \beta x^2 + (1-\beta)\gamma^2 \sqrt{1 + x^2/\gamma^2}, \qquad \text{for } 0 < \beta << 1 \text{ and } \gamma > 0.$$

The flow driven regularization term, proposed in [103], can be expressed as

$$E_{\text{reg}}(\vec{v}) = \sum_{\mathfrak{D}} \text{tr}(\varphi(\nabla v_x \nabla v_x^{\text{T}} + \nabla v_y \nabla v_y^{\text{T}}), \tag{2.28}$$

where tr is the trace operator, and the term in the argument of φ is called the structure tensor of $\nabla\vec{v}$. Computing the trace in the preceding equation is equivalent to finding the sum of eigenvalues of the structure tensor. Since the structure tensor is a symmetric positive semi-definite matrix, it has two orthonormal eigenvectors which are used to change the weighting of the regularizer across flow boundaries. Flow driven methods are computationally more expensive than image driven methods. Moreover, latter methods do not localize optical flow boundaries well compared to image driven methods.

Robust penalty functions have also been applied to regularization terms of variational methods. The use of either of the robust penalty functions given in Figure 2.5 for optical flow regularization leads to more penalization of small deviations from the smoothness constraint, and vice versa. Consequently, discontinuities in optical flow are better handled by these penalty functions than by the quadratic penalty. Modern state-of-the-art variational methods use the robust ℓ_1 norm with the gradient magnitude of optical flow [3, 4, 82, 30]. The ℓ_1 norm of the gradient magnitude is referred to as total variation (TV). It is worth mentioning that TV minimization favors a piecewise smooth optical flow with a sparse gradient field. We discuss the use of the ℓ_1 norm for the design of a novel sparsity promoting regularizer in Chapter 3.

Minimizing a continuous energy functional in variational methods involves the solution of partial differential equations. In these methods, the application of the gradient descent scheme generates a system of diffusion-reaction equations. Optical flow computed from

these equations is expected to preserve sharp motion boundaries when these equations are generated from regularizers using image or flow driven weighting schemes. In [103], a detailed discussion of partial differential equations and convex regularizers is provided for the computation of optical flow.

2.6 Occlusion Detection using Optical Flow

In any video sequence, it is most likely that moving objects will obstruct the view of other objects. The phenomenon of occlusion causes the blockage of some scene points and reveals some new portions of the scene across video frames. Occlusions play an important role in the visual perception of the three-dimensional shape of objects and the scene structure [104]. Occlusions are usually observed at moving object boundaries and discontinuities in the scene depth.

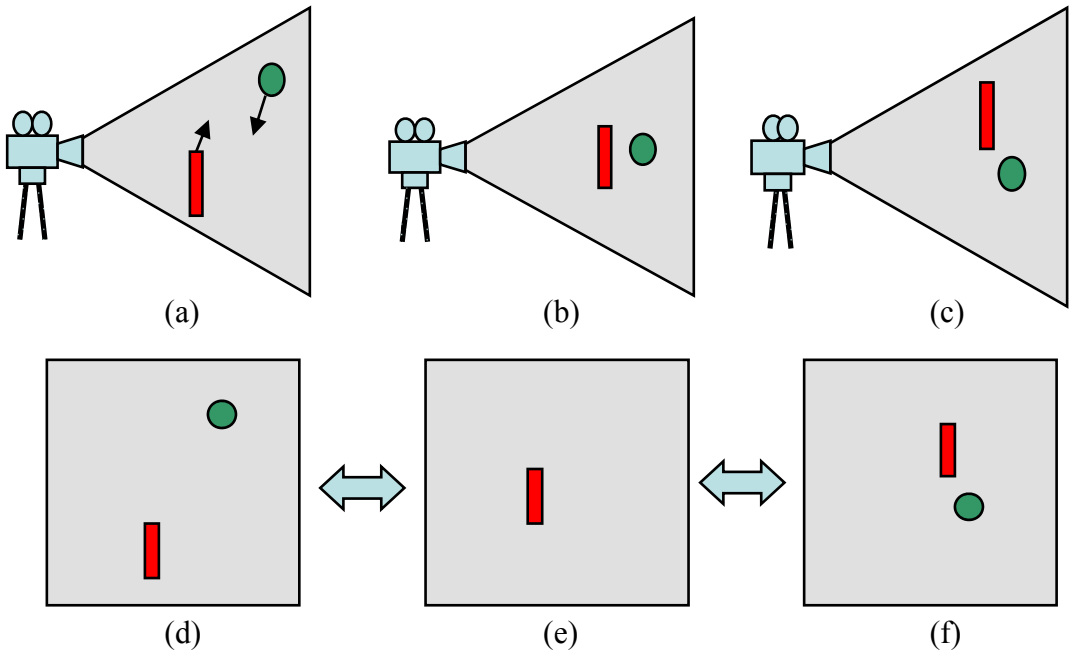


Figure 2.6: The phenomenon of occlusion in digital videos. A camera is capturing a scene at time (a) $t - 2dt$, (b) $t - dt$ and (c) t. The corresponding two-dimensional images are shown in (d)-(f), respectively. There is an occlusion between $t - 2dt$ and $t - dt$ whereas new scene points are revealed between $t - dt$ and t. Both situations cause problems for optical flow methods.

The matching of pixels across video frames is possible when objects in the scene remain visible over time. Pixels in a frame, not co-visible in the succeeding frame of the same sequence, should be taken as occluded pixels. Figure 2.6 shows the phenomenon of occlusion and dis-occlusion in digital videos. Occluded or dis-occluded regions of the scene may either be mismatched or not matched at all when constraints mentioned in Section 2.5.1 are used to form a data term. The data term should be trusted less in occluded regions; therefore, occlusion detection methods generally use a measure of the data term confidence for occlusion detection. The variational optical flow model with occlusion detection can be expressed as

$$E(\vec{v}) = \min_{\vec{v}} w(x,y) E_{\text{data}}(\vec{v}) + \lambda E_{\text{reg}}(\vec{v}), \tag{2.29}$$

where $w(x,y)$ is a positive weight that depends upon the occlusion detection confidence $o(x,y)$; it can be given as

$$w(x,y) = \max(1 - o(x,y), \delta).$$

A small positive constant δ ensures that the data term is always greater than 0. The larger is the $o(x,y)$, the less is the weight given to the data term.

Optical flow based occlusion detection methods use techniques from computational stereo to detect occluded pixels. Occlusion detection has been widely studied in two frame stereo vision [105, 106, 107, 108, 109]. Some stereo methods exploit the symmetry between left and right views of a stereo sequence to detect occlusion [105, 106, 109]. Since occlusions often occur at image discontinuities, some optical flow methods use image or flow driven anisotropic diffusion to detect occluded video regions by exploiting the motion symmetry [100, 110, 111]. These types of methods compute forward as well as backward optical flow, and use a consistency function between both flows. Video regions with low consistency are taken as occluded ones. The cross-check of the motion symmetry can be effective in occlusion detection; however, these methods compute optical flow bi-directionally, which is computationally expensive. A method which does not compute optical flow bidirectionally is presented in [82]. This method is based on an observation

that pixels in the reference video frame should uniquely correspond to pixels in the target frame [107]. The mapping of multiple pixels in the target frame to the same pixel in the reference frame is used as an indication of occluded pixels in this method. For the occlusion detection confidence $o(x, y)$, the mapping uniqueness criteria can be given as

$$o(x, y) = F(n(x - v_x, y - v_y) - 1, 0, 1),$$

where n represents the number of pixels mapped to the position $(x - v_x, y - v_y)$ in the target video frame. The function F truncates this number if it is outside the $[0, 1]$ range.

The residual of the OFC can also be used to find occluded regions when the mapping of scene points is not possible across video sequences. In [112], the residual of the OFC is used in a bilateral filter to detect occlusion. This method weights the data term according to the magnitude of the OFC residual. Since it is desirable to use a differentiable function for the weight function $w(x, y)$, this method uses

$$w(x, y) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left[\Big(I(x, y, t) - I(x + v_x, y + v_y, t - dt) \Big)^2 - \epsilon \right], \tag{2.30}$$

where the soft threshold parameter ϵ is used to decide whether a pixel is occluded or not. Another method uses a similar kind of approach for occlusion detection after representing the motion as particles [113]. The residual of a data term, which represents particles to be matched, is combined with the divergence of optical flow to detect occlusion. In [114], a neural network approach is proposed in parallel to optical flow for occlusion detection. The difference of velocities computed by optical flow and a neural network reveals occluded video regions. This approach is then used to track occluded objects. In [115], occlusions are detected from video frames by first dividing each video frame into numerous triangular shapes which are allowed to move relative to each other. The resulting occlusion effects are then detected using a numerical scheme. This approach is computationally expensive because it uses multiple video frames to detect occlusion.

Occlusion detection has also been considered in probabilistic optical flow estimation methods [116, 117, 118, 119]. In [116], a bidirectional Bayesian framework is used to estimate optical flow. The consistency of the computed motion is then used to reveal

occluded video regions. In [117], the histogram of pixel intensities is used, along with a consistency term based on nonlinear diffusion, to detect occluded pixels. In [118, 119], video frames are represented as overlapping moving layers in a probabilistic graphical model. This method segments moving layers, and uses a consistency function for the layer segmentation process to detect occluded pixels.

2.7 Conclusions

This chapter provides a description of commonly used optical flow estimation methods. It is intended to highlight strengths and weaknesses of these methods. Vision applications which use specific class of optical flow estimation model, according to the required accuracy and the computational complexity, have also been highlighted. This chapter does not use any strict classification of optical flow. Global methods, as an example, have been presented as parametric or nonparametric but the same could be applied to local gradient based methods. Global nonparametric variational optical flow methods have been discussed in detail because the subsequent work proposed in this thesis is based on these methods.

Chapter 3

A Sparsity Enhancing Regularizer for Variational Methods

Chapter contents

3.1 Introduction

Most of image processing and computer vision tasks are inverse problems where the aim is to find the solution of an unknown signal from inadequate and possibly noisy observations. Variational methods stabilize the solution of these ill-posed problems by regularizing unknown signals [120, 121]. The regularization is required to obtain physically plausible solutions. A good regularizer ensures a stable estimation of the unknown signal; therefore, the design of an efficient variational regularizer is crucial.

Variational methods impose a smoothness constraint to regularize ill-posed image processing tasks. A quadratic regularizer blurs strong edges and object boundaries by penalizing intensity variations at or across them, as mentioned in Section 2.5.2. To protect sharp edges and boundaries, robust norms are used with the smoothness constraint [59]. The ℓ_1 norm is of particular interest because it makes the variational functional convex. The noteworthy total variation (TV) regularizer of Rudin et al. uses the ℓ_1 norm with the smoothness term [122]. TV regularization has been successfully used in numerous computer vision and image processing tasks such as image denoising [122, 123, 124, 125], image restoration [126, 127, 128], image deconvolution [129, 130], image deblurring [131, 125], optical flow estimation [2, 132, 3, 133, 134, 82, 101], and stereo vision [135, 136, 137]. The TV regularizer promotes the sparsity of the computed solution. It preserves object boundaries better than a quadratic regularizer; however, it performs poor in low-textured image regions because it generates undesirable staircase artefacts. Thus, this chapter proposes a novel sparsity enhancing regularizer, which aims to overcome shortcomings of the TV regularizer. The proposed regularizer is applied to the problems of image denoising and image restoration from incomplete measurements. Extensive experimental results are conducted to reveal that image restoration using the proposed regularizer generally results in higher PSNR of the restored images than local TV regularizers.

The rest of the chapter is organized as follows. Section 3.2 discusses sparsity promoting regularization for variational methods. Isotropic, anisotropic and higher order total variation regularization are focussed in this discussion. Section 3.3 proposes a novel

sparsity enhancing regularizer that enforces the continuity of partial derivatives of the unknown signal. The rotational invariance of the proposed regularizer is proved. An analysis is conducted with the TV regularizer to analyze sparsity promoting capabilities of the proposed regularizer. Section 3.4 embeds the proposed regularizer into a variational framework to solve two common problems in image restoration: image denoising and image reconstruction from incomplete measurements. It also gives algorithmic details to solve aforementioned image restoration problems. Section 3.5 presents experimental results to show the superiority of the proposed regularizer over total variation for noisy image restoration. It also demonstrates the capability of the proposed regularizer to reconstruct images from highly incomplete measurements. Section 3.6 concludes the chapter.

3.2 Sparsity promoting TV Regularization

TV regularization has been introduced to the task of edge preserving image denoising by Rudin, Osher and Fatemi (ROF) [122]. TV and its variants have been extensively used in image processing and computer vision tasks ever since. This section presents sparsity promoting isotropic, anisotropic and higher order TV regularizers.

3.2.1 Isotropic Total Variation

An isotropic quantity does not change its value regardless of its direction of measurement. An isotropic regularizer applies the same amount of regularization in each direction [122]. Let a digital image $F(i,j)$ be defined for the horizontal and vertical co-ordinates i and j, respectively, over a domain $\mathfrak{D}$. The discrete isotropic TV (iTV) of $F(i,j)$ can be defined as the sum of the magnitude of the image gradient at each pixel:

$$||\nabla F(i,j)||_{\text{iTV}} := \sum_{i,j\in\mathfrak{D}} |\nabla F(i,j)| = \sum_{i,j\in\mathfrak{D}} \sqrt{[\nabla_x F(i,j)]^2 + [\nabla_y F(i,j)]^2}. \tag{3.1}$$

The sum of the gradient magnitude makes TV a semi ℓ_1 norm. It is well-known that ℓ_1 norm of the gradient promotes the sparsity of the image in the gradient domain. Therefore, a piecewise smooth image is obtained by minimizing (3.1) An isotropic TV can preserve

sharp horizontal and vertical edges; however, it causes the unnecessary smoothing of those edges and strong intensity regions which are at an angle other than 0° or 90°. This problem arises because isotropic TV minimizes the gradient magnitude. This problem can be reduced by using variants of TV, for example, anisotropic TV [138, 139, 81, 140], nonlocal TV [141, 142] or higher order TV [143, 144].

3.2.2 Anisotropic Total Variation

Anisotropic TV applies a direction dependent regularization to the underlying image. It imposes the smoothing along strong intensity structures but not across them [145]. For a discrete image $F(i,j)$, the anisotropic TV (aTV) can be defined as the sum of the absolute difference of partial image derivatives:

$$||\nabla F(i,j)||_{\text{aTV}} := \sum_{i,j\in\mathfrak{D}} |\nabla_x F(i,j)| + |\nabla_y F(i,j)|. \tag{3.2}$$

Anisotropic TV regularization performs better than isotropic TV at strong intensity structures such as edges and object boundaries. However, unlike isotropic TV, it is not rotationally invariant to the pixel grid. Thus, it produces suboptimal solutions in the presence of rotations of the camera or the pixel grid [146].

Discontinuities in an image occur along object boundaries where the image gradient is high. Therefore, making the regularization adaptive to the image structure can preserve sharp boundaries better than a non-adaptive regularization. To this end, anisotropic TV regularization is sometimes weighted by an image-driven weight function $w(|\nabla I|)$ as

$$E_{\text{reg}}(F(i,j)) = \sum_{i,j\in\mathfrak{D}} w(|\nabla I|)\Big(|\nabla_x F(i,j)| + |\nabla_y F(i,j)|\Big). \tag{3.3}$$

For small positive numbers α and β, $w(|\nabla I|)$ can be chosen as $w(|\nabla I|) = \exp(-\alpha|\nabla I|^{\beta})$. Anisotropic TV is easier to minimize than its isotropic counterpart. Therefore, numerous convex minimization methods can be used to minimize anisotropic TV. These include gradient methods [147, 148, 149, 150], primal dual methods [127, 151], iterative shrinkage or thresholding based methods [152, 153, 125], and graph cuts based methods [154, 155, 146].

3.2.3 Higher Order Total Variation

Isotropic and anisotropic TV produce staircase artefacts in low-textured and flat image regions. To reduce this undesirable effect, higher order total variation regularization has been proposed [143]. Mathematically, higher order total variation (HOTV) can be given as

$$||\nabla F(i,j)||_{\text{HOTV}} := \sum_{i,j\in\mathfrak{D}} |\nabla F(i,j)| + \alpha|\nabla^k F(i,j)|, \tag{3.4}$$

where k represents the order of the regularization, and a positive constant α balances the effect of gradient and higher order derivatives. The idea of using higher order derivatives has been modified to include Laplacian Δ with the gradient for the regularization of unknown images as

$$||\nabla F(i,j)||_{\text{Lap-TV}} := \sum_{i,j\in\mathfrak{D}} |\nabla F(i,j)| + \alpha|\Delta F(i,j)|. \tag{3.5}$$

The use of higher order derivatives may result in the blurring of sharp image boundaries. Thus, higher order TV regularization uses an adaptive functional which makes the regularizer act as ordinary TV at sharp boundaries, whereas it uses higher order derivatives in textured and flat image regions.

Total generalized variation (TGV) has been proposed as a generalization of higher order TV regularization [144, 156]. By changing the order of the regularizer, TGV allows to reconstruct piecewise smooth, affine and quadratic images. The TGV regularizer can be used to obtain a globally optimal solution because, similar to the TV regularizer, it is also convex. TGV and HOTV are computationally more expensive than the ordinary TV because of the calculation of higher order derivatives.

3.3 Proposed Sparsity Enhancing Regularizer

This section proposes a novel sparsity enhancing regularizer capable of avoiding the shortcomings of TV based regularizers. The proposed regularizer is based on the variational measure introduced in [157], which imposes the intensity continuity of partial image

derivatives at each pixel in a small neighborhood. The variational measure is defined for scalar images only. However, the proposed regularizer is designed to handle multi-channel vector valued images. This regularizer, in contrast to TV based regularizers, can preserve edges and object boundaries which are not either horizontal or vertical. It is also able to reduce undesirable staircase artefacts produced by TV in flat regions where there is a little intensity change. First, the regularizer is formulated for multi-channel images. Second, the rotational invariance of the proposed regularizer is proved for multi-channel images. Finally, its sparsity enhancing capabilities are theoretically compared against the local TV regularizer.

3.3.1 The Formulation

Let $\mathbf{F} = [F_1 \quad F_2 \quad \cdots F_c]^{\mathrm{T}}$ be a multi-channel image with c number of channels. Discrete partial derivatives of this image $\mathbf{F}$ at pixel location (i, j) can be given as forward differences:

$$\nabla_{x(i,j)}\mathbf{F}(i,j) = \mathbf{F}(i+1,j) - \mathbf{F}(i,j),$$

and

$$\nabla_{y(i,j)}\mathbf{F}(i,j) = \mathbf{F}(i,j+1) - \mathbf{F}(i,j).$$

Now, for each channel of $\mathbf{F}$, let us consider the continuity of its partial derivatives in a 2×2 neighborhood. Partial derivatives $\nabla_{x(i,j)}$ and $\nabla_{y(i,j)}$ can be continuous along all directions except their own directions because they are desired to be discontinuous to preserve sharp edges and boundaries. For example, $\nabla_{x(i,j)}$ enforces the continuity along all directions except the horizontal direction. The continuity of partial derivatives depends upon the direction associated with the boundary. Continuity constraints for different boundary directions in a 2×2 neighborhood are as follows:

- vertical, $\nabla_{xx(i,j)} = \nabla_{x(i,j+1)} - \nabla_{x(i,j)} = 0;$
- horizontal, $\nabla_{yy(i,j)} = \nabla_{y(i+1,j)} - \nabla_{y(i,j)} = 0;$
- diagonal 45°, $\nabla_{xy(i,j)} = \nabla_{x(i,j)} + \nabla_{y(i+1,j)} = 0;$

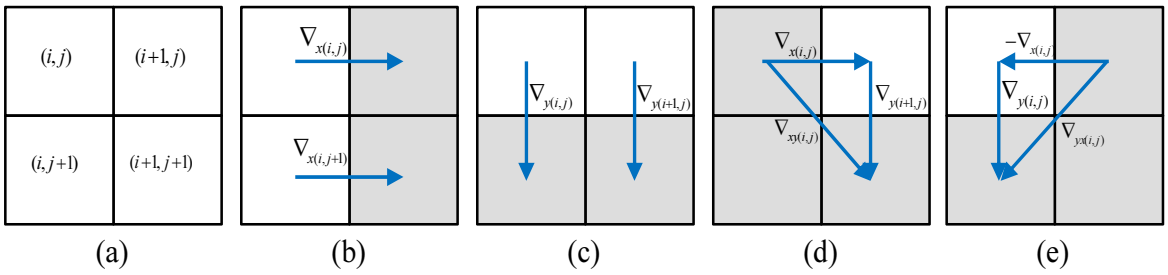


Figure 3.1: The gradient continuity: (a) a 2×2 neighborhood showing pixel positions, (b)-(e) the vertical, horizontal, diagonal 45° and diagonal 135° boundaries, respectively. Required derivatives for the gradient continuity are shown in blue colour in (b)-(e), for each direction of the boundary.

- diagonal 135°, $\quad \nabla_{yx(i,j)} = \nabla_{y(i,j)} - \nabla_{x(i,j)} \quad = 0.$

Figure 3.1 shows these four boundaries along with associated image derivatives in a 2×2 neighborhood. For vertical, horizontal, diagonal 45° and 135° boundaries, the directional continuity of partial derivatives ∇_x and ∇_y is enforced by minimizing $\nabla_{xx(i,j)}$, $\nabla_{yy(i,j)}$, $\nabla_{xy(i,j)}$ and $\nabla_{yx(i,j)}$, respectively. To regularize the multi-channel image $\mathbf{F}$, we minimize the ℓ_1 norm of aforementioned partial derivatives as

$$\begin{aligned} E_{\text{reg}}(\mathbf{F}) = &||(\nabla_x \mathbf{F})||_1^2 + ||(\nabla_y \mathbf{F})||_1^2 + ||(\nabla_{xy} \mathbf{F})||_1^2 \\ &+ ||(\nabla_{yx} \mathbf{F})||_1^2 + ||(\nabla_{xx} \mathbf{F})||_1^2 + ||(\nabla_{yy} \mathbf{F})||_1^2. \end{aligned} \tag{3.6}$$

A careful inspection of continuity constraints reveals that

$$\nabla_{xx}\mathbf{F} = \mathbf{F}(i+1, j+1) + \mathbf{F}(i, j) - \mathbf{F}(i+1, j) - \mathbf{F}(i, j+1) = \nabla_{yy}\mathbf{F}.$$

Given the goal to recover sparsest partial derivatives, $\nabla_{xx(i,j)} = 0$ or $\nabla_{yy(i,j)} = 0$ implies zero partial derivative along the horizontal or vertical direction, respectively. This is equivalent to $\nabla_{x(i,j)} = 0$ or $\nabla_{y(i,j)} = 0$. In this case, the minimization of $||\nabla_{xx}\mathbf{F}||_1$ and $||\nabla_{yy}\mathbf{F}||_1$ is redundant under the minimization of either $||\nabla_x \mathbf{F}||_1$ or $||\nabla_y \mathbf{F}||_1$. Therefore, these two terms can be omitted in (3.6). Since $\mathbf{F} = [F_1 \quad F_2 \quad \cdots F_c]^{\mathrm{T}}$, we penalize the magnitude of horizontal, vertical and diagonal derivatives of each channel, and denote

this regularizer as HVD($\mathbf{F}$):

$$E_{\text{reg}}(\mathbf{F}) = \text{HVD}(\mathbf{F}) :=$$
$$||\sqrt{(\nabla_x F_1)^2 + (\nabla_x F_2)^2 + \cdots + (\nabla_x F_c)^2}||_1^2 + ||\sqrt{(\nabla_y F_1)^2 + (\nabla_y F_2)^2 + \cdots + (\nabla_y F_c)^2}||_1^2 +$$
$$||\sqrt{(\nabla_{xy} F_1)^2 + (\nabla_{xy} F_2)^2 + \cdots + (\nabla_{xy} F_c)^2}||_1^2 + ||\sqrt{(\nabla_{yx} F_1)^2 + (\nabla_{yx} F_2)^2 + \cdots + (\nabla_{yx} F_c)^2}||_1^2. \quad (3.7)$$

The regularizer HVD($\mathbf{F}$) enforces the continuity of partial image derivatives at each pixel, and minimizes their ℓ_1 norm separately. Therefore, regularizing an image using (3.7) is expected to preserve sharp horizontal, vertical as well as diagonal edges. Furthermore, the inclusion of diagonal derivatives along with the horizontal and vertical derivatives in a neighborhood around each pixel imposes more constraints on the image to be restored. Consequently, the HVD regularizer reduces staircase artefacts in flat regions. In addition, it is more robust against outliers than traditional TV regularizer. The separate minimization of partial derivatives favors a solution which is sparser than the solution obtained by using TV. As an implication, HVD($\mathbf{F}$) requires fewer number of measurements than TV for the estimation of unknown signals.

3.3.2 Rotational Invariance

We prove the rotational invariance of the proposed regularizer given in (3.7) for multi-channel images. We use 2D rotations to give proof for 2-channel images; nevertheless, by using higher dimensional rotations, it is easy to show that the regularizer is invariant to rotations for multi-channel images. Let $\mathbf{R}$ be a 2D rotation matrix for a 2-channel image $\mathbf{F} = [F_1 \ F_2]^{\text{T}}$. When the camera is rotated by an an angle θ, the rotated image $\mathbf{RF}$ is given as

$$\mathbf{RF} = \left(\begin{array}{c|c} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{array} \right) \left(\begin{array}{c} F_1 \\ \hline F_2 \end{array} \right) \quad (3.8)$$

or

$$\mathbf{RF} = \mathbf{R}_1 F_1 + \mathbf{R}_2 F_2, \quad (3.9)$$

where $\mathbf{R} = \left(\begin{array}{c|c} \mathbf{R}_1 & \mathbf{R}_2 \end{array} \right)$. For the proposed regularizer, we will prove that HVD($\mathbf{F}$) = HVD($\mathbf{RF}$). Considering the first term in the square root of (3.7), i.e., $(\nabla_x F_1)^2 + (\nabla_x F_2)^2$,

and substituting rotated image $\mathbf{RF}$ into this term, we get

$$\begin{aligned}(\nabla_x \mathbf{R}_1 F_1)^2 + (\nabla_x \mathbf{R}_2 F_2)^2 &= (\nabla_x F_1 \cos\theta - \nabla_x F_2 \sin\theta)^2 + (\nabla_x F_1 \sin\theta + \nabla_x F_2 \cos\theta)^2, \\ &= (\nabla_x F_1 \cos\theta)^2 + (\nabla_x F_2 \sin\theta)^2 - 2(\nabla_x F_1 \cos\theta)(\nabla_x F_2 \sin\theta) \\ &+ (\nabla_x F_1 \sin\theta)^2 + (\nabla_x F_2 \cos\theta)^2 + 2(\nabla_x F_1 \sin\theta)(\nabla_x F_2 \cos\theta).\end{aligned}$$

By canceling common terms and after some rearrangements, we obtain

$$\begin{aligned}(\nabla_x \mathbf{R}_1 F_1)^2 + (\nabla_x \mathbf{R}_2 F_2)^2 &= (\nabla_x F_1)^2(\cos^2\theta + \sin^2\theta) + (\nabla_x F_2)^2(\cos^2\theta + \sin^2\theta), \\ &= (\nabla_x F_1)^2 + (\nabla_x F_2)^2,\end{aligned}$$

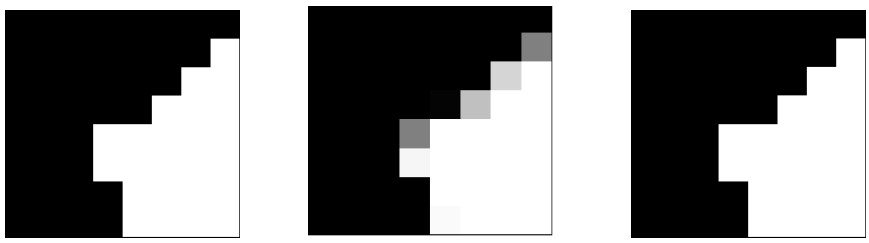
which is identical to the image without rotation. The similar proof can be provided for terms involving ∇_y, ∇_{xy} and ∇_{yx}. Hence the proposed regularizer is invariant to camera rotations.

3.3.3 Sparsity Comparison with Total Variation

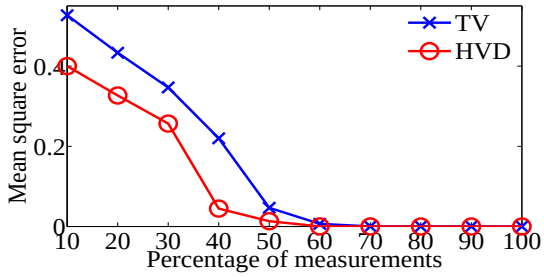
We compare the sparsity promoting capabilities of the TV and proposed HVD regularizers. Let us denote the set of partial derivatives as $z = \{x, y, xy, yx\}$. The magnitude of these derivatives can be computed as $|\nabla_z \mathbf{F}|$. The proposed regularizer penalizes the magnitude of each partial derivative separately; therefore, we compare the sparsity of each partial derivative with the sparsity of the gradient magnitude which is minimized in TV. The gradient magnitude is given as

$$|\nabla \mathbf{F}| = \sqrt{(\nabla_x \mathbf{F})^2 + (\nabla_y \mathbf{F})^2}.$$

The sparsity of the gradient magnitude $|\nabla \mathbf{F}|$ is the number of nonzero $\nabla_x \mathbf{F}$ or $\nabla_y \mathbf{F}$. Hence, $|\nabla \mathbf{F}|$ is less sparse than either $|\nabla_x \mathbf{F}|$ or $|\nabla_y \mathbf{F}|$. Diagonal derivatives depend upon the horizontal and vertical derivatives as shown in Figure 3.1. Therefore, their sparsity is also less than or equal to the sparsity of the gradient magnitude $|\nabla \mathbf{F}|$. Hence each partial derivative is either sparser than the gradient magnitude or its sparsity is equal to the sparsity of the gradient magnitude. Thus, a sparse signal can be reconstructed by



(a) (b) (c)



(d)

Figure 3.2: Diagonal edge recovery using TV and HVD: (a) a sharp diagonal edge, its reconstruction from 50% measurements using (b) TV and (c) HVD. Note that the edge is recovered correctly by HVD in (c). (d) The mean square error as a function of percentage of measurements.

TV as well as by the proposed regularizer because HVD($\mathbf{F}$) produces a solution that is sparser than the solution obtained by TV. A consequence of the above analysis is that the proposed regularizer requires fewer measurements than TV for the recovery of an unknown image from highly incomplete measurements.

To illustrate the sparsity comparison of HVD and TV, we recover a diagonal image by both regularizers. Figure 3.2 shows the recovery of a diagonal image edge of size 8×8 pixels from reduced measurements. An undesirable blurring of the edge can be seen in TV based recovery, whereas HVD correctly recovers the sharp edge. Figure 3.2 (d) shows the recovery of the same edge from different percentage of measurements by using TV and HVD. It is obvious that HVD requires fewer measurements than TV for a mean square error close to zero.

3.3.4 Reduction in Staircase Artefacts

Local TV regularizers produce staircase artefacts in low-textured image regions because these regularizers under-constrain the underlying image in flat regions. To mitigate undesirable staircase artefacts, low-textured regions should be constrained more. The proposed HVD regularizer includes diagonal derivatives along with the horizontal and vertical derivatives in a neighborhood around each pixel. Thus, it constrains the image more than local TV regularizers. Consequently, staircase artefacts are considerably reduced by the proposed regularizer. Another advantage of using diagonal derivatives is that HVD preserves diagonal edges better than TV regularizers. In addition, the separate minimization of partial derivatives makes HVD more robust against outliers than traditional TV regularizers which penalize the horizontal and vertical flow derivatives only.

3.4 Image Restoration using Proposed Regularizer

Restoration techniques aim to recover or reconstruct images degraded by some known degradation phenomena. The problem of image restoration models the degradation process and applies the inverse of this process to reconstruct the original image. In this section, we apply the proposed regularizer to the problem of image restoration. Let $\mathbf{f} = [\mathbf{f}_1 \quad \mathbf{f}_2 \quad \cdots \mathbf{f}_c]^{\mathrm{T}}$ be the lexicographically vectorized multi-channel image $\mathbf{F}$, $\mathbf{g}$ be the degraded version of $\mathbf{f}$ and S be a linear matrix operator that represents the degradation process. The image restoration model can now be given as

$$\mathbf{g} = S\mathbf{f} + \eta, \tag{3.10}$$

where η denotes multichannel noise. One popular example of the restoration process is image denoising. When the matrix S is assumed to be an identity matrix $\mathcal{I}_n$ of size $n \times n$, we get $\mathbf{g}$ to be a noisy version of the original image $\mathbf{f}$. We denoise (3.10) using the proposed regularizer. The variational energy $E(\mathbf{f})$ incorporating (3.10) and the proposed regularizer can now be given as

$$E(\mathbf{f}) = \min_{\mathbf{f}} ||\mathbf{f} - \mathbf{g}||_1^2 + \lambda[\mathrm{HVD}(\mathbf{f})], \tag{3.11}$$

where λ is a regularization parameter. The first term in (3.11) represents the data fidelity term that enforces the restored image to be close to the noisy image. Note that we have used the robust ℓ_1 norm with the data fidelity term to handle outliers in the restoration process.

A second example of the restoration process is image reconstruction from incomplete measurements. In this case, the matrix S is an underdetermined linear operator which measures the image $\mathbf{f}$ such that the dimension of the measured image $\mathbf{g}$ is less than the original image $\mathbf{f}$. The image reconstruction using the proposed regularizer can be given as

$$E(\mathbf{f}) = \min_{\mathbf{f}} ||S\mathbf{f} - \mathbf{g}||_1^2 + \lambda[\mathrm{HVD}(\mathbf{f})]. \tag{3.12}$$

Note that both data and regularization terms in (3.12) are convex; thus, convex optimization methods can be used to solve for $\mathbf{f}$ from the resulting energy $E(\mathbf{f})$. Here, we demonstrate how a fast algorithm, NESTA, presented in [150] can be modified to solve Equations (3.11) and (3.12). NESTA has been used to solve large-scale variational problems [158]. Here, we give the algorithmic details for the image reconstruction. The same algorithm can be applied to the image denoising problem by taking the matrix S as an identity matrix. NESTA uses a differentiable Huber norm approximation to the ℓ_1 norm; therefore, it can handle smooth as well as nonsmooth convex functionals. We modify NESTA to solve image restoration problem. The Huber norm is given as

$$||x||_\epsilon = \begin{cases} \frac{x^2}{2\epsilon}, & \text{if } |x| \leq \epsilon, \\ |x| - \frac{\epsilon}{2}, & \text{otherwise.} \end{cases}$$

The derivative of the Huber norm is given by

$$\frac{\partial}{\partial x}||x||_\epsilon = \frac{x}{\max(\epsilon, |x|)}. \tag{3.13}$$

We use differentiable Huber norm in place of the ℓ_1 norm in (3.12). The combined data

and the regularization energy $E(\mathbf{f})$ is now given as

$$E(\mathbf{f}) = \min_{\mathbf{f}} ||S\mathbf{f} - \mathbf{g}||_{\epsilon}^{2} + \lambda\Bigg(|| \sqrt{(\boldsymbol{\nabla}_x \mathbf{f}_1)^2 + (\boldsymbol{\nabla}_x \mathbf{f}_2)^2 + \cdots + (\boldsymbol{\nabla}_x \mathbf{f}_c)^2} ||_{\epsilon}^{2} + || \sqrt{(\boldsymbol{\nabla}_y \mathbf{f}_1)^2 + (\boldsymbol{\nabla}_y \mathbf{f}_2)^2 + \cdots + (\boldsymbol{\nabla}_y \mathbf{f}_c)^2} ||_{\epsilon}^{2} + || \sqrt{(\boldsymbol{\nabla}_{xy} \mathbf{f}_1)^2 + (\boldsymbol{\nabla}_{xy} \mathbf{f}_2)^2 + \cdots + (\boldsymbol{\nabla}_{xy} \mathbf{f}_c)^2} ||_{\epsilon}^{2} + || \sqrt{(\boldsymbol{\nabla}_{yx} \mathbf{f}_1)^2 + (\boldsymbol{\nabla}_{yx} \mathbf{f}_2)^2 + \cdots + (\boldsymbol{\nabla}_{yx} \mathbf{f}_c)^2} ||_{\epsilon}^{2} \Bigg), \tag{3.14}$$

where $\boldsymbol{\nabla}_x$, $\boldsymbol{\nabla}_y$, $\boldsymbol{\nabla}_{xy}$ and $\boldsymbol{\nabla}_{yx}$ are sparse difference matrices to calculate derivatives of vectorized images. An iterative scheme is used to find the minimum of (3.14) at iteration k as

$$\frac{\partial E(\mathbf{f}^k)}{\partial \mathbf{f}^k} = \partial_{\mathbf{f}^k} E(\mathbf{f}^k) = \Bigg[2S^{\mathrm{T}}(\mathbf{f}^k - \mathbf{g}) + \lambda \Bigg(\boldsymbol{\nabla}_x^{\mathrm{T}} \frac{(\boldsymbol{\nabla}_x \mathbf{f}^k)}{\max(\epsilon, |\boldsymbol{\nabla}_x \mathbf{f}^k|)} + \boldsymbol{\nabla}_y^{\mathrm{T}} \frac{(\boldsymbol{\nabla}_y \mathbf{f}^k)}{\max(\epsilon, |\boldsymbol{\nabla}_y \mathbf{f}^k|)} + \boldsymbol{\nabla}_{xy}^{\mathrm{T}} \frac{(\boldsymbol{\nabla}_{xy} \mathbf{f}^k)}{\max(\epsilon, |\boldsymbol{\nabla}_{xy} \mathbf{f}^k|)} + \boldsymbol{\nabla}_{yx}^{\mathrm{T}} \frac{(\boldsymbol{\nabla}_{yx} \mathbf{f}^k)}{\max(\epsilon, |\boldsymbol{\nabla}_{yx} \mathbf{f}^k|)} \Bigg) \Bigg]. \tag{3.15}$$

The algorithm computes two auxiliary variables $\mathbf{p}^k$ and $\mathbf{q}^k$ at each iteration from $\partial_{\mathbf{f}^k} E(\mathbf{f}^k)$. It then combines both auxiliary variables to get next estimate $\mathbf{f}^k$. The choice of ϵ plays an important role in the algorithm. The speed of the convergence is shown to have direct relationship with this approximation constant [150]. A small value of ϵ gives good accuracy at the cost of slow convergence and vice versa. The proposed algorithm uses Lipschtz continuity; therefore, a Lipschitz constant L is required for the computation of auxiliary variables from (3.15). Lipschitz constant L depends on λ, Huber norm parameter ϵ, and the norms of sparse difference matrices.

To compute the Lipschitz constant L, we need to find the upper bound for the norm of HVD. It has been shown in [159] that difference matrices used to calculate TV are bounded above by 8. A similar analysis can be made for difference matrices used in HVD. The ℓ_1 norm of any matrix is maximum absolute column sum of that matrix. HVD consists of four sparse difference matrices: $\boldsymbol{\nabla}_x$, $\boldsymbol{\nabla}_y$, $\boldsymbol{\nabla}_{xy}$ and $\boldsymbol{\nabla}_{yx}$, which are used to compute discrete differences. These matrices have exactly two nonzero entries $+1$ or -1. Therefore, they satisfy $||\boldsymbol{\nabla}_x||_1 = 2$, $||\boldsymbol{\nabla}_y||_1 = 2$, $||\boldsymbol{\nabla}_{xy}||_1 = 2$ and $||\boldsymbol{\nabla}_{yx}||_1 = 2$. Since we

minimize the magnitude of each partial flow derivative in HVD,

$$||\boldsymbol{\nabla}_x^{\mathrm{T}}\boldsymbol{\nabla}_x||_1 + ||\boldsymbol{\nabla}_y^{\mathrm{T}}\boldsymbol{\nabla}_y||_1 + ||\boldsymbol{\nabla}_{xy}^{\mathrm{T}}\boldsymbol{\nabla}_{xy}||_1 + ||\boldsymbol{\nabla}_{yx}^{\mathrm{T}}\boldsymbol{\nabla}_{yx}||_1 = 4+4+4+4 = 16. \tag{3.16}$$

Hence, difference matrices used in HVD are bounded above by 16. Lipschitz constant L is then given as $L = 16\lambda/\epsilon$. The algorithm runs for a fixed number of iterations or until it reaches the convergence. The summary of the algorithm is shown in Table 3.1.

Table 3.1: The proposed algorithm for image restoration.

Initialization: $\mathbf{f}^0 = 0$.
Set iteration index $k = 1$,
while(not converged & $k \leq$ max_iter)
1. compute $\partial_{\mathbf{f}^k} E(\mathbf{f}^k)$ from (3.15),
2. compute $\gamma^k = \frac{1}{2}(k+1)$ and $\tau^k = \frac{2}{k+3}$,
3. compute $\mathbf{p}^k = \mathbf{f}^k - \frac{1}{L}\partial_{\mathbf{f}^k} E(\mathbf{f}^k)$,
4. compute $\mathbf{q}^k = \mathbf{f}^0 - \frac{1}{L}\sum_i^k \gamma^i, \partial_{\mathbf{f}^i} E(\mathbf{f}^i)$,
5. update $\mathbf{f}^k = \tau^k \mathbf{p}^k + (1-\tau^k)\mathbf{q}^k$,

$k = k+1$,
end while

3.5 Experimental Results

The proposed regularizer has been tested on several images to evaluate its performance under noise and incomplete measurements. For the validation of the proposed regularizer, a comparative analysis is conducted with isotropic, anisotropic and higher order TV regularizers by assessing the quality of the restored images. First, the experimental setup, describing images used in experiments, is presented. Second, the restoration performance is analyzed on images corrupted by a controlled amount of noise. Third, images are restored from randomly sensed incomplete measurements. Finally, the computational complexity of the proposed method, and the sensitivity of the regularization parameter is assessed.

Table 3.2: Optimal regularization parameters for different regularizers.

λ_{iTV}	λ_{aTV}	λ_{HVD}	λ_{HOTV}
0.05	0.05	0.01	0.006

3.5.1 Experimental Setup

All experiments have been performed on publically available real world images which are used as benchmarks for various vision and image processing tasks. The image dataset used in these experiments comprises of greyscale images *Cameraman*, *Barbara*, *Boat* and *Man*, and colour images *Baboon*, *House*, *Monarch* and *Pepper*. These images are corrupted by a controlled amount of noise. The quality of restored images have been assessed by calculating the peak signal to noise ratio (PSNR), which is given as

$$\text{PSNR} = 20 * \log(\frac{f_{\max}^2}{\text{MSE}}), \tag{3.17}$$

where $f_{\max} = 255$ for an 8 bit image and MSE is the mean squared error:

$$\text{MSE} = \frac{1}{n}\sum_n(\mathbf{f} - \mathbf{g}^2). \tag{3.18}$$

To conduct a fair comparison of aforementioned three TV regularizers and the proposed regularizer, same algorithm has been used to minimize their variational regularization energies. Therefore, the energies of proposed regularizer and the three TV regularizers are minimized using NESTA ([150]). It should be mentioned that the use of different regularizers alter the variational energy to be minimized. Consequently, the values of optimum regularization parameters for these regularizers also change. In these experiments, we have manually tuned regularization parameters of these regularizers to get best restoration results for all of these regularizers. Table 3.2 shows the values of these parameters for our experiments.

3.5.2 Image Denoising

These experiments have been conducted to test the capability of the proposed regularizer to denoise images. A controlled amount of Gaussian noise is added to images described

Table 3.3: PSNR results on all eight images for optimum values of regularization parameters.

Image	Anisotropic TV	Isotropic TV	Proposed HVD	Higher order TV
Cameraman	27.85	27.18	**30.21**	29.38
Barbara	28.65	28.27	30.28	**30.43**
Boat	26.6	26.94	**30.46**	30.07
Man	27.36	27.02	**31.09**	30.42
Baboon	26.95	26.15	29.84	**30.67**
House	27.61	27.09	**30.48**	30.12
Monarch	28.58	28.07	**30.92**	30.15
Pepper	28.64	28.56	**30.67**	29.81

above. Since clean images are available, performances of the proposed and TV based regularizers have been measured quantitatively by calculating the PSNR for denoised images. The influence of the controlled noise is also analyzed qualitatively on denoised images.

Table 3.3 shows quantitative results for all four regularizers on eight images. These results have been taken for a standard deviation of noise $\sigma = 25$. Optimum values of regularization parameters are used for all regularizers. It can be observed that the HOTV regularizer performs better than both isotropic and anisotropic TV regularizers. However, the proposed HVD regularizer outperforms the HOTV for most of images.

Figure 3.3 demonstrates the denoising of the greyscale image *Cameraman* when it is contaminated with a noise of $\sigma = 25$. Highlighted parts of images in Figure 3.3 (c) and (d) show staircase artefacts, whereas highlighted parts in Figure 3.3 (e) and (f) show a significant reduction in these artefacts. However, the image denoised by the HVD has a higher value of PSNR = 30.21 as compared to the image denoised by the HOTV regularizer with a PSNR = 29.38, as given in Table 3.3. Highlighted parts of Figure 3.3 are also shown in Figure 3.4 as 3D plots for a better visualization of staircase artefacts.

Denoising results on the colour image *Monarch* are shown in Figure 3.5. A qualitative comparison of images in Figure 3.5 (b) and (c) with (d) and (e) reveals that HVD and HOTV regularizers outperform anisotropic and isotropic TV especially in highlighted

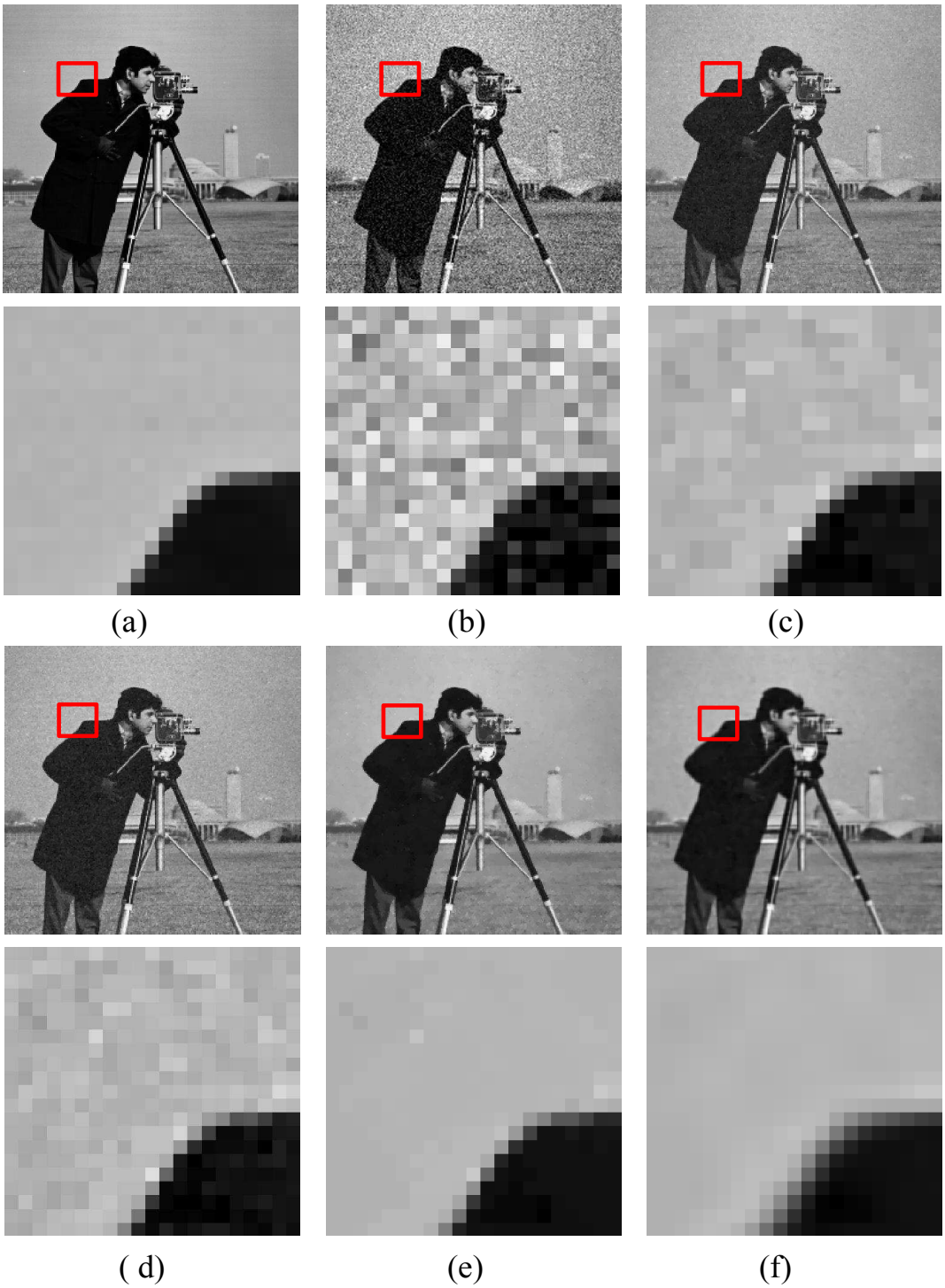


Figure 3.3: The denoising of the greyscale image *Cameraman*. (a) Original image, (b) image corrputed by a Gaussian noise of $\sigma = 25$, image denoised by (c) anisotropic TV, (d) isotropic TV, (e) the proposed HVD and (f) higher order TV regularizers. Highlighted parts of these images are also given in the bottom.

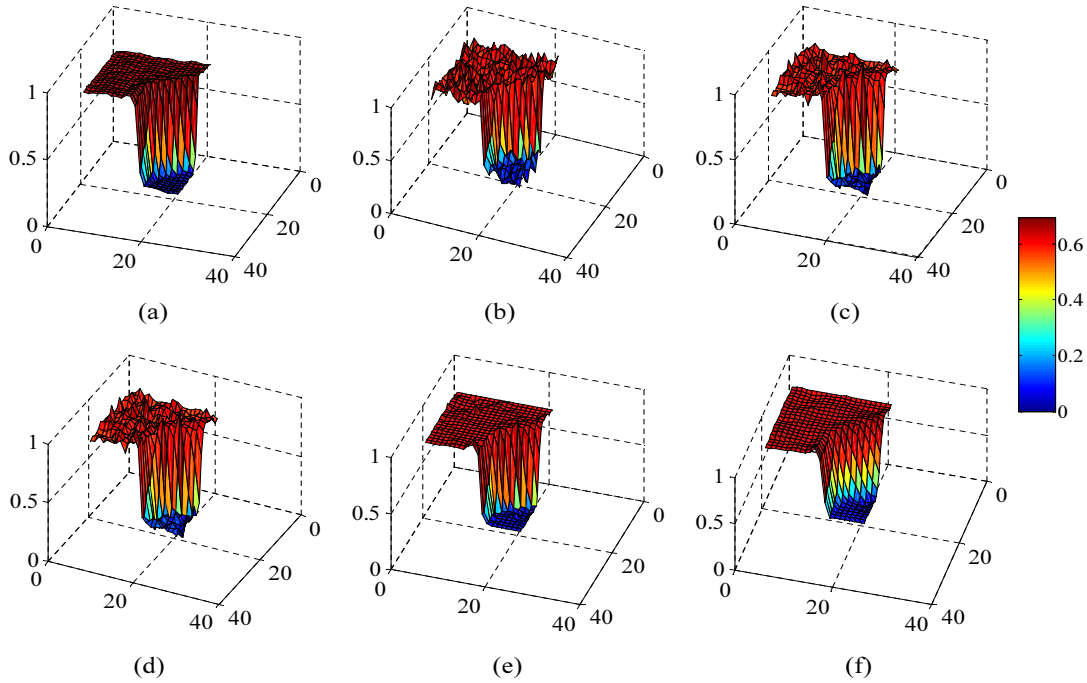


Figure 3.4: Highlighted parts of Figure 3.3 shown as 3D plots for the visualization of staircase artefacts.

textured regions of denoised images. The image denoised by the HVD regularizer in Figure 3.5 (d) has a higher PSNR = 30.92 than anisotropic, isotropic and higher order TV regularizers with PSNR = 28.58, 28.07 and 30.15, respectively. Moreover, staircase artefacts can be observed in highlighted parts of images in Figure 3.5 (b) and (c). HVD and HOTV regularizers do not show noticeable staircase artefacts.

Figure 3.6 presents PSNR as a function of standard deviation of noise σ for all four regularizers. An average PSNR has been calculated over all eight images, and the results are reported in Figure 3.6 (a). These results clearly indicate that the proposed regularizer outperforms TV based regularizers for increasing values of standard deviation of noise. Similar kind of results can be observed for *Cameraman* and *Pepper* in Figure 3.6 (b) and (c), respectively.

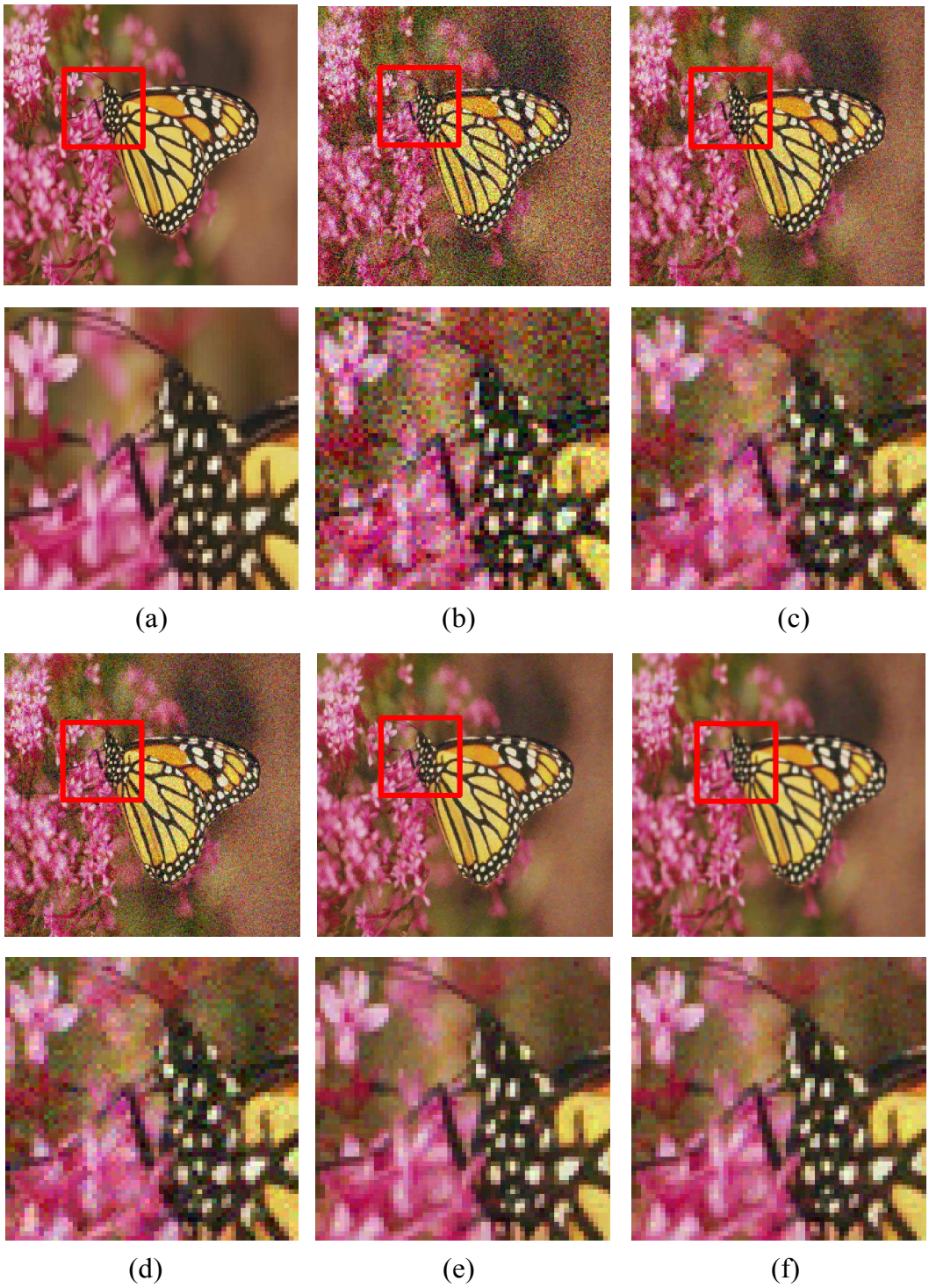
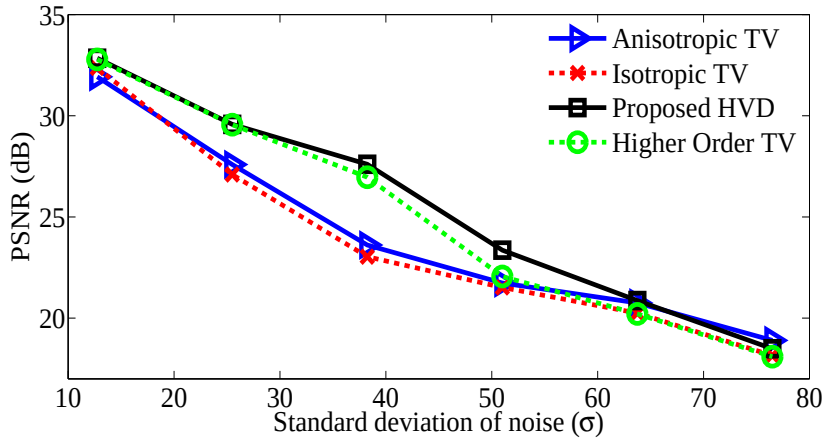
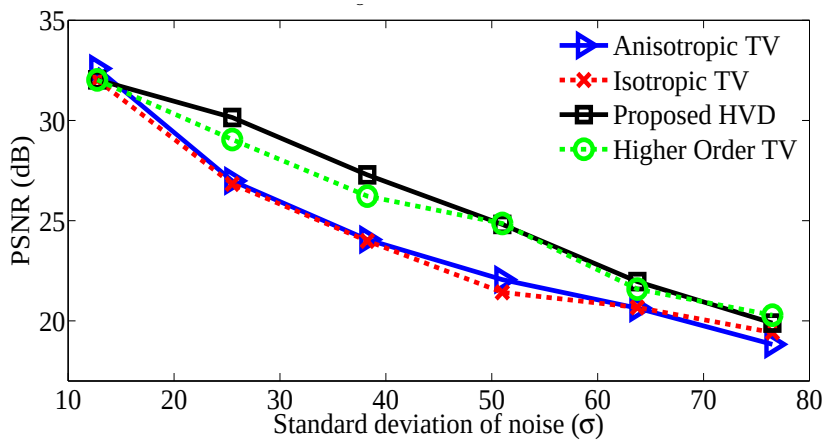


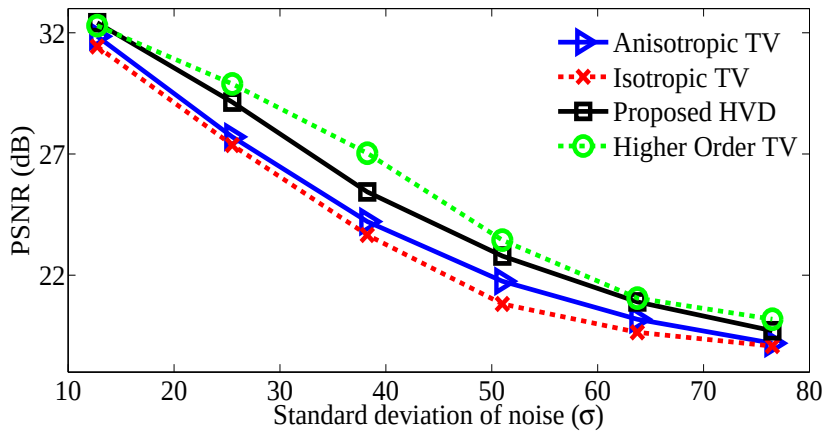
Figure 3.5: The denoising of the colour image *Monarch*. (a) Original image, (b) image corrputed by a Gaussian noise of $\sigma = 25$, image denoised by (c) anisotropic TV, (d) isotropic TV, (e) the proposed HVD and (f) higher order TV regularizers. Highlighted parts of these images are also given in the bottom.



(a)



(b)



(c)

Figure 3.6: The PSNR as a function of the standard deviation σ of noise for anisotropic TV, isotropic TV, proposed HVD and higher order TV regularizers. Results for (a) the whole dataset of 8 images (b) *Cameraman* and (c) *Peppers*. Regularization parameters of all four regularizers have been tuned to get best PSNR performances.

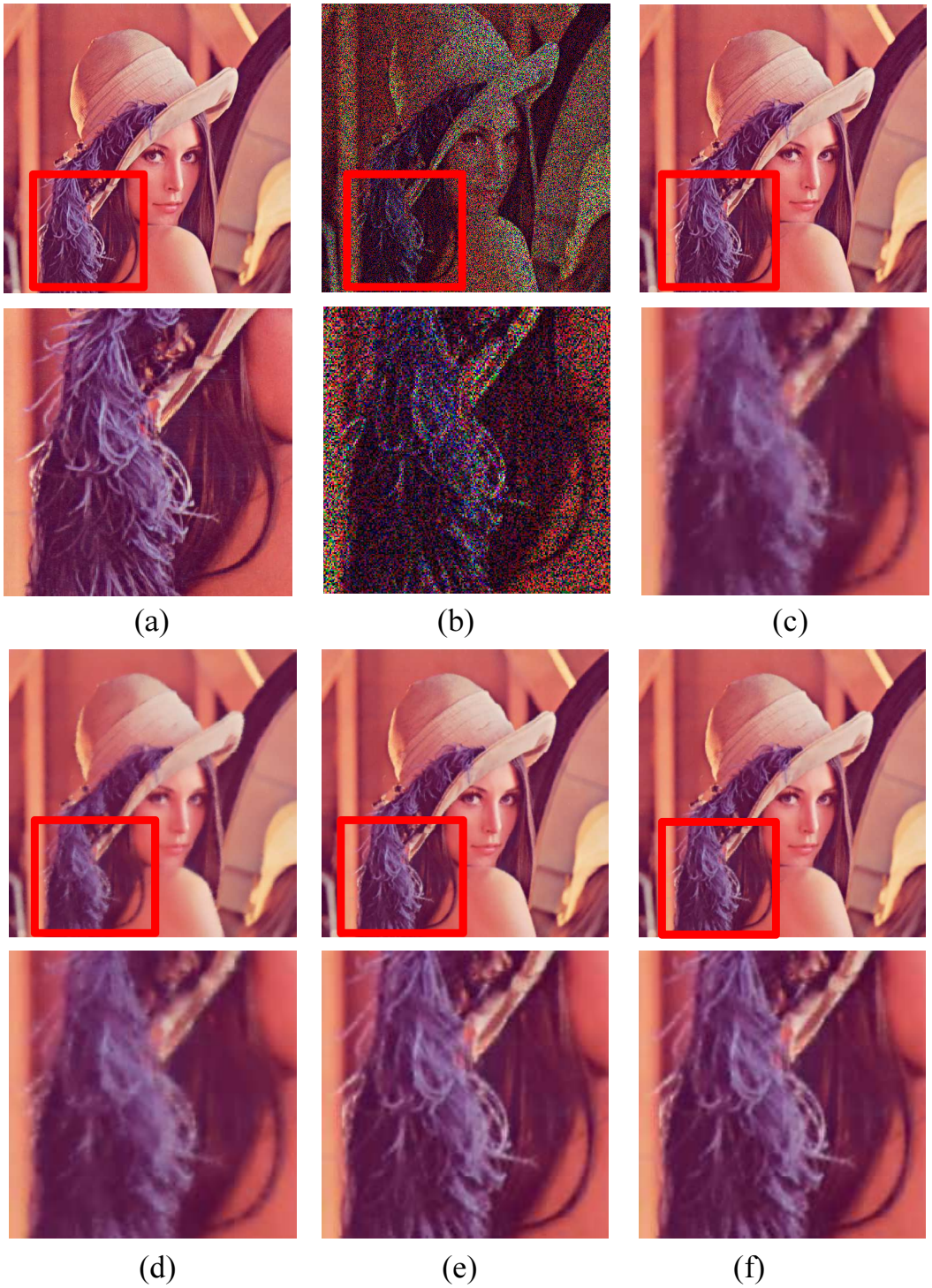


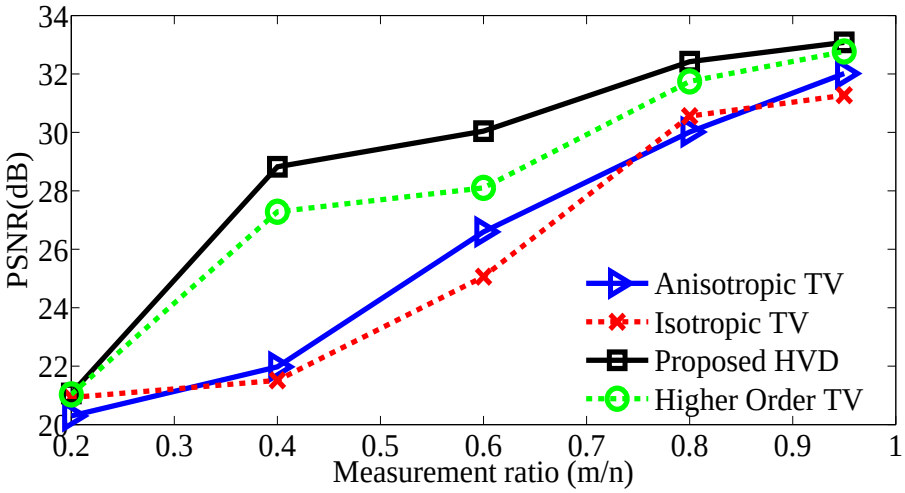
Figure 3.7: Reconstruction of the colour image *Lena* from 50% measurements. (a) Original image, (b) 50% randomly selected pixels of the original image, image reconstructed by (c) anisotropic TV, (d) isotropic TV, (e) the proposed HVD and (f) higher order TV regularizers. Highlighted parts of these images are also given in the bottom.

3.5.3 Reconstruction using Incomplete Measurements

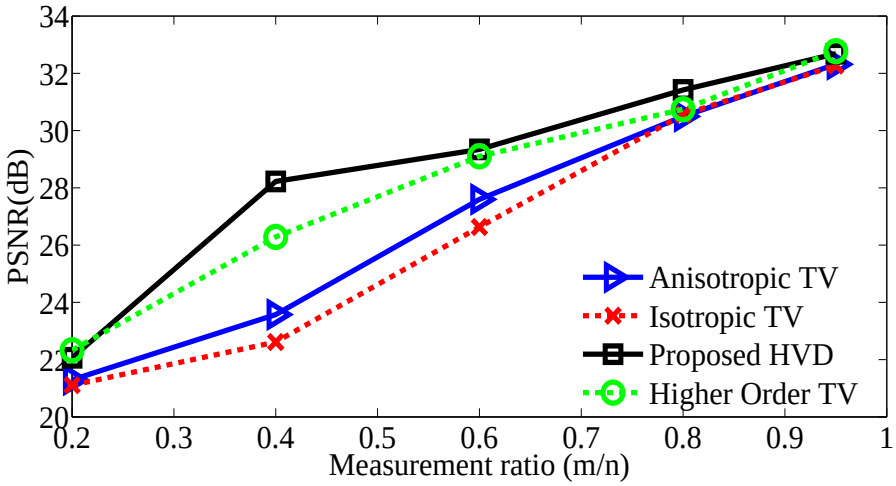
In these experiments, the performance of the proposed regularizer is tested against TV based regularizers for the reconstruction of images from incomplete measurements. We define m/n as measurement ratio for m number of randomly made measurements in an n pixel image. The performance of each regularizer is evaluated for different measurement ratios. Note that if $m/n = 1$, then full measurements are used, i.e., there is no reduction. Measurements are gradually increased, and the PSNR is plotted against each set of measurement ratio. Since image measurements are made randomly, we run all regularizers multiple times for each measurement ratio. In these experiments, we take the average of results after we run each regularizer five times to reconstruct images from randomly chosen measurements.

Figure 3.7 reports the reconstruction of the colour image *Lena* for a measurement ratio of 0.5. Results in highlighted parts of Figure 3.7 (c) and (d) show that TV regularizers blur highly textured hair of Lena. A significant decrease of this blurring can be observed by the proposed and higher order TV regularizers as shown in Figure 3.7 (e) and (f), respectively. However, the image reconstructed by HVD in Figure 3.7 (e) has a higher PSNR = 29.47 than the higher order TV with PSNR = 27.95.

Figure 3.8 presents the PSNR as a function of measurement ratio m/n. Results in Figure 3.8 (a) show the superiority of the proposed HVD over TV based regularizers for the reconstruction of *Lena* from small measurement ratios. Results in Figure 3.8 (b) are reported by averaging the PSNR of 8 reconstructed images against each set of measurement ratio. It can be observed that the proposed HVD regularizer is able to reconstruct images from measurements smaller than measurements required by TV based regularizers. Higher order TV regularizer performs better than isotropic and anisotropic regularizers; however, the proposed regularizer shows even superior results in terms of the PSNR of reconstructed images.



(a)



(b)

Figure 3.8: The PSNR as a function of the measurement ratio (m/n). These results are plotted for (a) *Lena* and (b) the whole dataset of 8 images.

3.5.4 Computational Complexity and Parameter Sensitivity

In this subsection, we perform experiments to compute the complexity and parameter sensitivity of all four regularizers. Images are corrupted by a Gaussian noise of $\sigma = 25$. For a fair comparison, noisy images are restored by TV based regularizers by using the same algorithm that has been used for the proposed regularizer.

Figure 3.9 shows time taken by all four regularizers to restore noisy image *Pepper* of

different sizes. Note that a logarithmic scale is used on y-axis for time consumed. The plot reveals that isotropic and anisotropic TV regularizers have nearly similar complexity. The proposed and higher order TV regularizers are computationally more expensive than other two regularizers. For small images, the difference between the computational complexity of the proposed regularizer and the higher order TV is not significant. However, it can be noticed that the proposed regularizer is considerably faster than the higher order TV for large images.

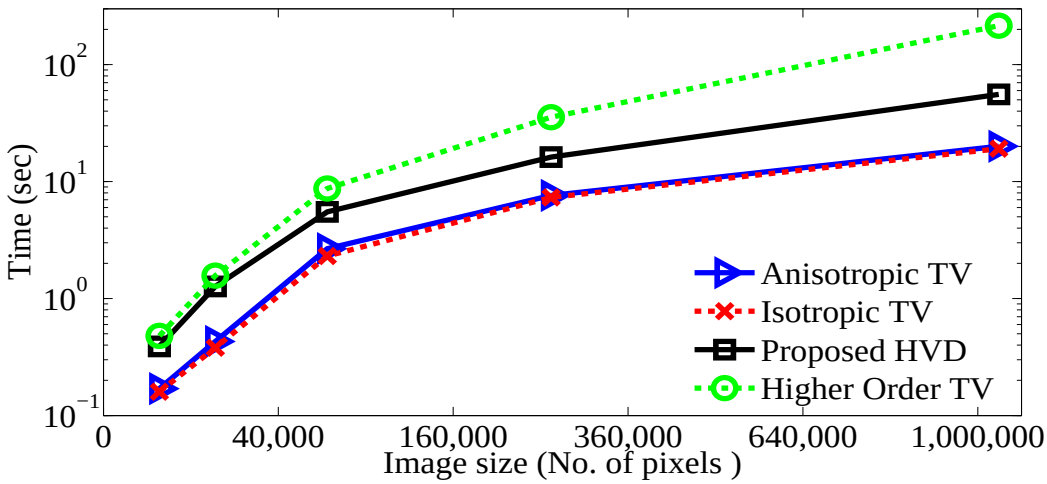


Figure 3.9: The computational complexity of all four regularizers on images of different sizes.

Figure 3.10 shows the effect of changing regularization parameters λ on PSNR performances of regularizers. Variational regularizers are known to produce good image restoration results for a range of regularization parameter values. The larger the range, the more robust is the regularizer against change in the value of λ. In these results, it can be observed that the proposed regularizer shows more robustness to the change of the parameter than other regularizers.

3.6 Conclusions

This chapter was dedicated to the regularization of ill-posed imaging problems. We investigated sparsity enhancing regularizers in the context of variational methods. A novel

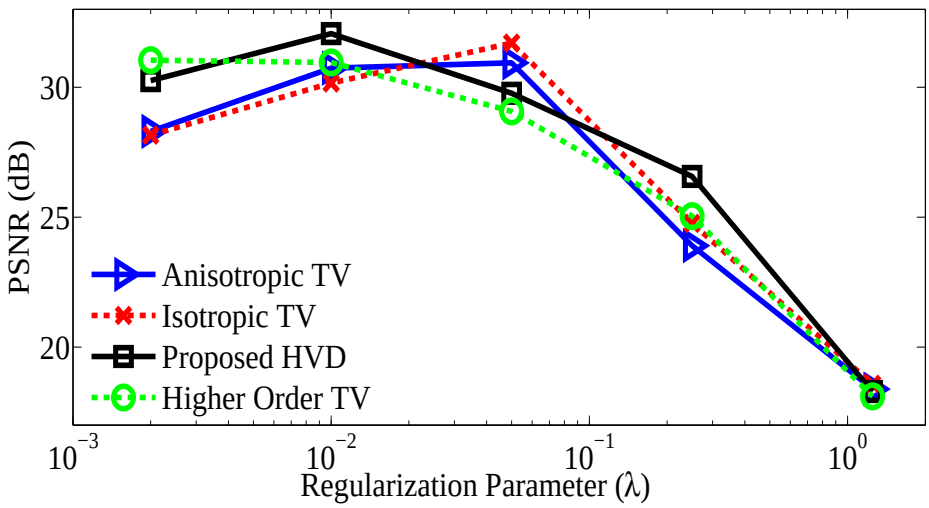


Figure 3.10: The sensitivity analysis of TV and proposed regularizers against the regularization parameter λ. Eight images have been used for this analysis.

variational regularizer HVD was presented after we discussed shortcomings of existing TV based regularizers. The HVD regularizer was proven to be rotationally invariant to camera motions. A sparsity comparison of the proposed regularizer was conducted with total variation regularizer. It was revealed that the proposed regularizer can produce more sparse solutions than TV regularizer. TV regularizer is known to generate staircase artefacts in the computed solution. However, the HVD regularizer was shown to reduce these artefacts significantly.

The proposed regularizer was applied to the problems of image denoising and image reconstruction from incomplete measurements. Experiments were conducted to show that the proposed regularizer can produce results better than TV based regularizers. The improvement in results is fundamental because the proposed regularizer will be used in subsequent chapters of this thesis. It will be used to regularize optical flow estimation in Chapter 4 and occlusion detection using optical flow in Chapter 5.

Chapter 4

Dense Optical Flow using Sparse Regularizers

Chapter contents

4.1 Introduction

The retrieval of motion from digital videos is a major task of computer vision systems. Optical flow methods compute an estimate of the motion between consecutive video frames. There is an interest in estimating accurate optical flow because numerous vision tasks require accurate motion for the extraction of semantic information from video sequences. Object detection and tracking, surveillance, motion compensated coding, depth measurement, robot navigation, video stabilization, 3D structure from motion are among some

of the applications that use optical flow. Therefore, it is essential to obtain good optical flow estimates.

Optical flow estimation is an ill-posed problem because of infamous aperture problem. Variational methods regularize optical flow field to resolve aperture problem. This chapter is dedicated to the use of variational regularizers for optical flow estimation. The attention has been focused on those regularizers which promote the sparsity of the solution. For a real positive number p such that $0 \leq p \leq 1$, the ℓ_p pseudo-norm gives rise to sparse and robust regularization. However, for $p < 1$, the regularization term does not define a norm, and the resulting regularization is not convex. The use of the ℓ_1 norm is of particular interest in this context because it results in convex regularizers. Therefore, efficient convex optimization techniques can be used to minimize them. One of the most commonly used example of such robust and convex regularizer is total variation (TV), which uses ℓ_1 norm with the gradient magnitude [122]. TV penalizes large signal coefficients less than the small signal coefficients and vice versa. Consequently, discontinuities in optical flow are better handled by this regularizer than a quadratic regularizer. TV regularization favors a piecewise smooth optical flow with a sparse gradient field. However, regularization of optical flow using the TV regularizer suffers from the shortcomings as mentioned in Chapter 3. In this chapter, we use the regularizer proposed in Chapter 3 for the estimation of optical flow.

This chapter is organized as follows. Section 4.2 provides the relevant work on sparse regularization, such as total variation (TV) regularization, for optical flow estimation. Section 4.3 embeds the proposed sparse regularizer of Chapter 3 with numerous optical flow data terms in a variational framework. It also presents details of a first order, fast and convex algorithm to compute optical flow. Section 4.4 reports the experimental setup, results and performance evaluation on various types of synthetic and real video sequences to show the efficacy of the proposed method.

4.2 Relevant Background

Variational methods regularize ill-posed problems by introducing smoothness constraints or bounds on the norms of vector spaces. The regularization can be considered as a tradeoff between data fitting and norm reduction of the solution. The regularization of optical flow proposed by Horn and Schunck's [22] is based on Tikhonov regularization [160]. This is a quadratic regularization which uses Euclidean distance or ℓ_2 norm. A quadratic regularization does not respect discontinuities in the flow field because it causes blurring of motion boundaries by imposing the same amount of smoothness everywhere in the video frame.

To preserve motion boundaries, the regularization can be made adaptive to motion discontinuities. It is commonly observed in digital videos that motion boundaries coincide with object boundaries. Therefore, large image gradients are produced at these boundaries. Based on this observation, Nagel [99] introduced the concept of oriented smoothness which weights the smoothness according to the orientation of motion boundaries. Nagel's method applies the smoothness along motion boundaries but inhibit it across them. To prevent undesirable blurring of motion boundaries, robust penalty functions have been used with the regularization term of the variational optical flow model (2.15). The quadratic regularization of optical flow has been replaced by the regularization which either uses nonconvex Geman-McClure [59, 80] or Lorentzian penalty functions [161, 162]. The regularization using these robust functions is able to preserve sharp motion boundaries. The use of these penalty functions with the regularization term of variational optical flow promotes the sparsity of optical flow in the gradient domain.

TV has been used to regularize optical flow in [91, 92]. TV favors a piecewise smooth optical flow with a sparse gradient field; it preserves motion boundaries better than by a quadratic regularizer. Homogeneous TV regularization has been used for optical flow in [2]. However, homogeneous TV can cause unnecessary smoothing of those edges and strong intensity regions which are not horizontal or vertical, i.e., boundaries at an angle other than 0° or 90°. Since most of motion boundaries coincide with image disconti-

nuities, making the regularization adaptive to the image structure can better preserve motion boundaries. To this end, image structure adaptive TV [3, 132, 163, 82] have been employed.

The isotropic and anisotropic TV regularizers compute local derivative features of an image. These regularizers assumes that the image is a set of connected objects with each object smooth inside. Thus, these regularizers face difficulties in preserving the texture and fine details of objects. Moreover, unavailability of significant derivatives deteriorates performances of isotropic and anisotropic TV regularizers in flat and low-textured video regions. Undesirable staircase artefacts are thus developed in low-textured video regions. These problems can be mitigated by combining image gradients in a small neighborhood. Nonlocal TV has been used to mitigate these problems [134, 156]. Nonlocal TV is considerably slower than local isotropic and anisotropic TV for large nonlocal window sizes.

The matching term of variational methods can not deal with outliers when it is based on quadratic penalization. Therfore, the robust Charbonnier penalty [76, 4], and sparsity promoting convex ℓ_1 norm have been used with data matching terms of variational optical flow methods [92, 132]. Nonconvex penalty functions have also been used to construct data terms [80, 162]. However, it has been pointed out in [4] that the use of convex ℓ_1 norm gives better results than the use of nonconvex penalties with regularization terms of variational optical flow methods. This is partly due to the fact that, in contrast to nonconvex methods, efficient convex optimization methods are available.

A data term based on the OFC is unable to cope with brightness change scenarios. Under the violation of intensity constancy, intensity derivatives are relatively less affected than the intensity itself. The data term based on the constancy of first-order image derivatives have been used in [91, 95, 82]. The constancy of higher-order intensity derivatives is also investigated to construct the data term of variational optical flow model in [78]. The computation of higher-order derivatives is difficult for noisy videos. Moreover, it is expensive to compute higher-order derivatives. Thus, the constancy of higher-order image

derivatives is rarely used for optical flow estimation.

4.3 Optical Flow using Proposed Regularizer

In this section, we embed the proposed regularizer of Chapter 3 into a variational framework. The resulting variational energy of the data and the regularization terms is minimized using a first-order convex algorithm. The details of the proposed algorithm are also given in this section.

4.3.1 The Formulation

Let $\mathbf{I_x}$, $\mathbf{I_y}$ and $\mathbf{I_t}$ be lexicographically vectorized horizontal, vertical and temporal intensity derivatives, respectively. Moreover, let $\mathbf{v}_x$ and $\mathbf{v}_y$, respectively, denote the vectorized horizontal and vertical flow components. The linearized version of the OFC given in (2.2) can now be written as

$$\begin{pmatrix}\text{diag}(\mathbf{I_x}) & \text{diag}(\mathbf{I_y})\end{pmatrix}\begin{pmatrix}\mathbf{v}_x\\ \mathbf{v}_y\end{pmatrix} = -\mathbf{I_t}, \tag{4.1}$$

where $\text{diag}(\mathbf{I_x})$, $\text{diag}(\mathbf{I_y})$ are diagonal matrices made up of horizontal and vertical intensity derivative vectors, respectively. Let us define

$$A = [\text{diag}(\mathbf{I_x}) \;\; \text{diag}(\mathbf{I_y})], \qquad \mathbf{v} = [\mathbf{v}_x \;\; \mathbf{v}_y]^{\mathrm{T}} \qquad \text{and} \qquad \mathbf{y}_1 = -\mathbf{I_t}.$$

The data term in the matrix-vector form can now be given as

$$E_{\text{data1}}(\mathbf{v}) = ||A\mathbf{v} - \mathbf{y}_1||_2^2. \tag{4.2}$$

The proposed regularizer given in Equation (3.6) is applied to the vectorized optical flow $\mathbf{v}$ as

$$E_{\text{reg}}(\mathbf{v}) = ||(\boldsymbol{\nabla}_x\mathbf{v})||_1 + ||(\boldsymbol{\nabla}_y\mathbf{v})||_1 + ||(\boldsymbol{\nabla}_{xy}\mathbf{v})||_1 + ||(\boldsymbol{\nabla}_{yx}\mathbf{v})||_1. \tag{4.3}$$

The combined variational energy $E(\mathbf{v})$ of the data and the regularization terms given in Equations (4.2) and (4.3), respectively, is now expressed as

$$E(\mathbf{v}) = \min_{\mathbf{v}} ||A\mathbf{v} - \mathbf{y}_1||_2^2 + \lambda\Big[||(\boldsymbol{\nabla}_x\mathbf{v})||_1 + ||(\boldsymbol{\nabla}_y\mathbf{v})||_1 + ||(\boldsymbol{\nabla}_{xy}\mathbf{v})||_1 + ||(\boldsymbol{\nabla}_{yx}\mathbf{v})||_1\Big]. \tag{4.4}$$

Since Equation (4.4) uses a data term that is based on OFC, it can not handle intensity change scenarios. A data term based on the constancy of the intensity gradient, for example, can be embedded with the proposed regularizer. The constancy of the intensity gradient is given as

$$\nabla I(x, y, t) - \nabla I(x - v_x dt, y - v_y dt, t - dt) = 0. \tag{4.5}$$

The first-order approximation of this assumption is usually used to construct a data term that is robust against intensity change. The linearization of (4.5) results

$$\begin{pmatrix} I_{xx} v_x + I_{xy} v_y \\ I_{yx} v_x + I_{yy} v_y \end{pmatrix} = - \begin{pmatrix} I_{xt} \\ I_{yt} \end{pmatrix}, \tag{4.6}$$

$$\begin{pmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{pmatrix} \begin{pmatrix} v_x \\ v_y \end{pmatrix} = - \begin{pmatrix} I_{xt} \\ I_{yt} \end{pmatrix}. \tag{4.7}$$

Let $\mathbf{I_{xx}}$, $\mathbf{I_{xy}}$, $\mathbf{I_{yx}}$ and $\mathbf{I_{yy}}$ be lexicographically vectorized second-order intensity derivatives. Moreover, let $\mathbf{I_{xt}}$, $\mathbf{I_{yt}}$ represent second-order spatiotemporal intensity derivatives. The linearized version of the intensity gradient constancy assumption can now be written as

$$\begin{pmatrix} \mathrm{diag}(\mathbf{I_{xx}}) & \mathrm{diag}(\mathbf{I_{xy}}) \\ \mathrm{diag}(\mathbf{I_{yx}}) & \mathrm{diag}(\mathbf{I_{yy}}) \end{pmatrix} \begin{pmatrix} \mathbf{v}_x \\ \mathbf{v}_y \end{pmatrix} = - \begin{pmatrix} \mathbf{I_{xt}} \\ \mathbf{I_{yt}} \end{pmatrix}, \tag{4.8}$$

where $\mathrm{diag}(\mathbf{I_{xx}})$, $\mathrm{diag}(\mathbf{I_{xy}})$, $\mathrm{diag}(\mathbf{I_{yx}})$ and $\mathrm{diag}(\mathbf{I_{yy}})$ are diagonal matrices made up of second-order intensity derivative vectors $(\mathbf{I_{xx}})$, $(\mathbf{I_{xy}})$, $(\mathbf{I_{yx}})$ and $(\mathbf{I_{yy}})$, respectively. Let us define

$$B = \begin{pmatrix} \mathrm{diag}(\mathbf{I_{xx}}) & \mathrm{diag}(\mathbf{I_{xy}}) \\ \mathrm{diag}(\mathbf{I_{yx}}) & \mathrm{diag}(\mathbf{I_{yy}}) \end{pmatrix}, \qquad \text{and} \qquad \mathbf{y}_2 = - \begin{pmatrix} \mathbf{I_{xt}} \\ \mathbf{I_{yt}} \end{pmatrix}.$$

Now the data term utilizing the intensity gradient constancy assumption can be constructed as

$$E_{\mathrm{data2}}(\mathbf{v}) = ||B\mathbf{v} - \mathbf{y}_2||_2^2. \tag{4.9}$$

Note that both (4.2) and (4.9) have similar form. However, $E_{\mathrm{data2}}(\mathbf{v})$ is computationally more expensive than $E_{\mathrm{data1}}(\mathbf{v})$ because it involves the solution of two equations. When the brightness change is not abrupt, it can be modeled as an affine function of the image intensity as

$$I(x, y, t) - I(x - v_x dt, y - v_y dt, t - dt) = d(x, y, t) I(x, y, t) + c(x, y, t), \tag{4.10}$$

where the multiplier $d(x, y, t)$ permits change in the image contrast and the offset $c(x, y, t)$ allows for the mean brightness change[98]. The linearized GDIM is expressed as

$$I_x v_x + I_y v_y + I_t = dI + c. \tag{4.11}$$

Let $\mathbf{I}$, $\mathbf{d}$ and $\mathbf{c}$ be lexicographically vectorized I, d and c, respectively. The vector form of (4.11) is given as

$$\begin{pmatrix} \text{diag}(\mathbf{I_x}) & \text{diag}(\mathbf{I_y}) & -\text{diag}(\mathbf{I}) & -\mathcal{I} \end{pmatrix} \begin{pmatrix} \mathbf{v}_x & \mathbf{v}_y & \mathbf{d} & \mathbf{c} \end{pmatrix}^{\mathrm{T}} = -\mathbf{I_t}, \tag{4.12}$$

where $\mathcal{I}$ is an identity matrix. Let $C = \begin{pmatrix} \text{diag}(\mathbf{I_x}) & \text{diag}(\mathbf{I_y}) & -\text{diag}(\mathbf{I}) & -\mathcal{I} \end{pmatrix}$ and $\mathbf{u} = \begin{pmatrix} \mathbf{v}_x & \mathbf{v}_y & \mathbf{d} & \mathbf{c} \end{pmatrix}^{\mathrm{T}}$. The data term based on (4.12) can now be given as

$$E_{\text{data3}}(\mathbf{v}, \mathbf{d}, \mathbf{c}) = ||C\mathbf{u} - \mathbf{y}_1||_2^2. \tag{4.13}$$

The data term $E_{\text{data3}}(\mathbf{v}, \mathbf{d}, \mathbf{c})$ involves brightness change parameters which have to be estimated along with $\mathbf{v}$. A smoothness constraint is applied to estimate brightness change parameters because they vary slowly in spatial domain.

4.3.2 Algorithmic Details

To estimate optical flow from Equation (4.4), we use a fast algorithm, NESTA, presented in [150]. This algorithm has been used to solve large-scale variational problems [158]. NESTA uses a differentiable Huber norm approximation to the ℓ_1 norm; therefore, it can handle smooth as well as nonsmooth convex functions. The Huber norm is given as

$$||x||_\epsilon = \begin{cases} \frac{x^2}{2\epsilon}, & \text{if } |x| \leq \epsilon, \\ |x| - \frac{\epsilon}{2}, & \text{otherwise.} \end{cases}$$

The derivative of the Huber norm is given by

$$\frac{\partial}{\partial x}||x||_\epsilon = \frac{x}{\max(\epsilon, |x|)}. \tag{4.14}$$

We use differentiable Huber norm in place of the ℓ_1 norm in (4.4). The combined data and the regularization energy $E(\mathbf{v})$ is now given as

$$E(\mathbf{v}) = \min_{\mathbf{v}} \Bigg[\lambda \Bigg(|| \sqrt{(\boldsymbol{\nabla}_x \mathbf{v}_x)^2 + (\boldsymbol{\nabla}_x \mathbf{v}_y)^2} ||_\epsilon + || \sqrt{(\boldsymbol{\nabla}_y \mathbf{v}_x)^2 + (\boldsymbol{\nabla}_y \mathbf{v}_y)^2} ||_\epsilon$$
$$+ || \sqrt{(\boldsymbol{\nabla}_{xy} \mathbf{v}_x)^2 + (\boldsymbol{\nabla}_{xy} \mathbf{v}_y)^2} ||_\epsilon + || \sqrt{(\boldsymbol{\nabla}_{yx} \mathbf{v}_x)^2 + (\boldsymbol{\nabla}_{yx} \mathbf{v}_y)^2} ||_\epsilon \Bigg) + || \Big[A\mathbf{v} - \mathbf{y} \Big] ||_2 \Bigg]. \tag{4.15}$$

The minimum of (4.15) at iteration k is computed by using an iterative scheme as

$$E_1(\mathbf{v}) = \min_{\mathbf{v}} \Bigg[\lambda\Bigg(||\sqrt{(\boldsymbol{\nabla}_x\mathbf{v}_x)^2 + (\boldsymbol{\nabla}_x\mathbf{v}_y)^2}||_\epsilon + ||\sqrt{(\boldsymbol{\nabla}_y\mathbf{v}_x)^2 + (\boldsymbol{\nabla}_y\mathbf{v}_y)^2}||_\epsilon$$
$$+||\sqrt{(\boldsymbol{\nabla}_{xy}\mathbf{v}_x)^2 + (\boldsymbol{\nabla}_{xy}\mathbf{v}_y)^2}||_\epsilon + ||\sqrt{(\boldsymbol{\nabla}_{yx}\mathbf{v}_x)^2 + (\boldsymbol{\nabla}_{yx}\mathbf{v}_y)^2}||_\epsilon\Bigg) + ||\Big[A\mathbf{v} - \mathbf{y}\Big]||_2\Bigg]. \quad (4.16)$$

The algorithm computes two auxiliary variables $\mathbf{p}^k$ and $\mathbf{q}^k$ at each iteration from $\partial_{\mathbf{v}^k}E(\mathbf{v}^k)$. The proposed algorithm uses Lipschtz continuity; therefore, a Lipschitz constant L is required for the computation of auxiliary variables in (4.16). Lipschitz constant L is given as $L = 16\lambda/\epsilon$ as computed in Section 3.4.

Table 4.1: The proposed algorithm for optical flow estimation.

Initialization: $\mathbf{v}^0 = [\mathbf{v}_x^0\ \mathbf{v}_y^0]^{\mathrm{T}} = 0$.

for level=pyramid_levels, ... , 1
  if (level= pyramid_levels)
    $\mathbf{v} = \mathbf{v}^0$,
  else
    $\mathbf{v}^0 =$ up-sample($\mathbf{v}$),
  end
  Set iteration index $k = 1$,
  while(not converged & $k \leq$ max_iter)
    1. compute $\partial_{\mathbf{v}^k}E(\mathbf{v}^k)$ from (4.16),
    2. compute $\gamma^k = \frac{1}{2}(k+1)$ and $\tau^k = \frac{2}{k+3}$,
    3. compute $\mathbf{p}^k = \mathbf{v}^k - \frac{1}{L}\partial_{\mathbf{v}^k}E(\mathbf{v}^k)$,
    4. compute $\mathbf{q}^k = \mathbf{v}^0 - \frac{1}{L}\sum_i^k \gamma^i, \partial_{\mathbf{v}^i}E(\mathbf{v}^i)$,
    5. update $\mathbf{v}^k = \tau^k\mathbf{p}^k + (1-\tau^k)\mathbf{q}^k$,
    $k = k + 1$,
  end while
  $\mathbf{v} = \mathbf{v} + \mathbf{v}^0$.
end for

The proposed algorithm is applied in a coarse-to-fine pyramid to handle large displacements of pixels. A Laplacian image sequence pyramid is constructed which contains consecutive video frames at different resolutions for a coarse-to-fine estimation of optical flow. Optical flow is initialized at 0 in the coarsest level. The estimated optical flow $\mathbf{v}$ at each level of the pyramid is propagated to the next fine level as initial guess $\mathbf{v}^0$. The

algorithm runs for a fixed number of iterations at each pyramid level or until it converges. The summary of the algorithm is shown in Table 4.1.

4.4 Experimental Results and Analysis

In this section, we evaluate the performance of the proposed method, and analyze experimental results. First, the experimental setup is described in Section 4.4.1. Second, the proposed regularizer is embedded with various data terms to assess its performance for different energy functionals in Section 4.4.2. Third, to determine the capability of the proposed method to estimate optical flow from reduced number of measurements, experiments are presented in Section 4.4.3. Finally, the performance of the proposed method is compared to a number of existing methods in Section 4.4.4.

4.4.1 Experimental Setup

The proposed method is run in a pyramid with a down-sampling factor of 0.70 to handle large displacements. It is found that a value of $\lambda \in [0.001\ 0.1]$ gives reasonable accuracy; therefore, we manually tuned this parameter in our experiments. The value of ϵ in Huber norm approximation of the ℓ_1 norm plays an important role in the convergence of the algorithm. In these experiments, a value of $\epsilon = 0.01$ is used which gives reasonable accuracy with a good convergence rate. The algorithm converges in a few hundred iterations; therefore, it is run for a maximum of 1000 iterations or until it converges.

The color scheme given in Figure 1.1 (e) has been used to visualize the magnitude and the direction of the estimated optical flow. Quantitative results are reported in terms of the mean end point error (MEPE) for video sequences with known ground-truth optical flow. The MEPE is given as

$$\text{MEPE} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\Big[(\mathbf{v}_x)_i - (\mathbf{v}_{x\text{GT}})_i\Big]^2 + \Big[(\mathbf{v}_y)_i - (\mathbf{v}_{y\text{GT}})_i\Big]^2}, \tag{4.17}$$

where $\mathbf{v}_{\text{GT}} = [\mathbf{v}_{x\text{GT}}\ \mathbf{v}_{y\text{GT}}]^{\text{T}}$ is the ground-truth flow, and n denotes the number of pixels in one video frame.

4.4.2 Proposed Regularizer with Various Data terms

Middlebury database is used for these experiments because it provides eight synthetically generated training video sequences with known ground-truth [164]. All sequences, except *Dimetrodon*, undergo motion of rigid objects or camera motion; thus, their ground-truth optical flow is piecewise smooth. *Dimetrodon* has a small part of the background undergoing a nonrigid motion. *Grove2*, *Grove3* and *Hydrangea* sequences have motion of thin structures like leaves, whereas *Rubberwhale* contains several independently moving objects. *Urban2* and *Urban3* have large motion of objects, and *Venus* is a stereo pair.

Figure 4.1 shows the estimated optical flows as color plots. It is clear from Figure 4.1 (b) and (c) that there is not much visible difference between the ground-truth and estimated optical flows. These results show that motion boundaries are preserved by the proposed method. Figure 4.1 (d) indicates that the error occurs at the occluded boundaries only. This error is prominent in *Urban2* and *Venus* sequences because these two sequences exhibit occlusions at image boundaries.

Table 4.2 presents quantitative results in term of MEPE for aforementioned video sequences. Five different data terms have been used with the proposed regularizer. Data terms which use the optical flow constraint (OFC) and the OFC based neighborhood dependent constraint (NDC) [132] perform well when brightness constancy assumption is followed. However, other three data terms, i.e., the gradient constancy assumption (GCA), the generalized dynamic image model (GDIM) [98] and the normalized optical flow constraint (N-OFC), can handle brightness change. *Rubberwhale* and *Urban3* sequences exhibit brightness change. GCA, GDIM and N-OFC perform better than other data terms on these two sequences in terms of the MEPE.

We have also applied the proposed method on real world sequences. A value of $\lambda = 0.005$ is found to be optimal for the sequences used in these experiments. To gain robustness against brightness change, real video sequences are pre-processed. A Gaussian filter of size 9×9 and a standard deviation of 1 is used for the smoothing of these sequences. These smoothed sequences are subtracted from raw sequences, and resultant

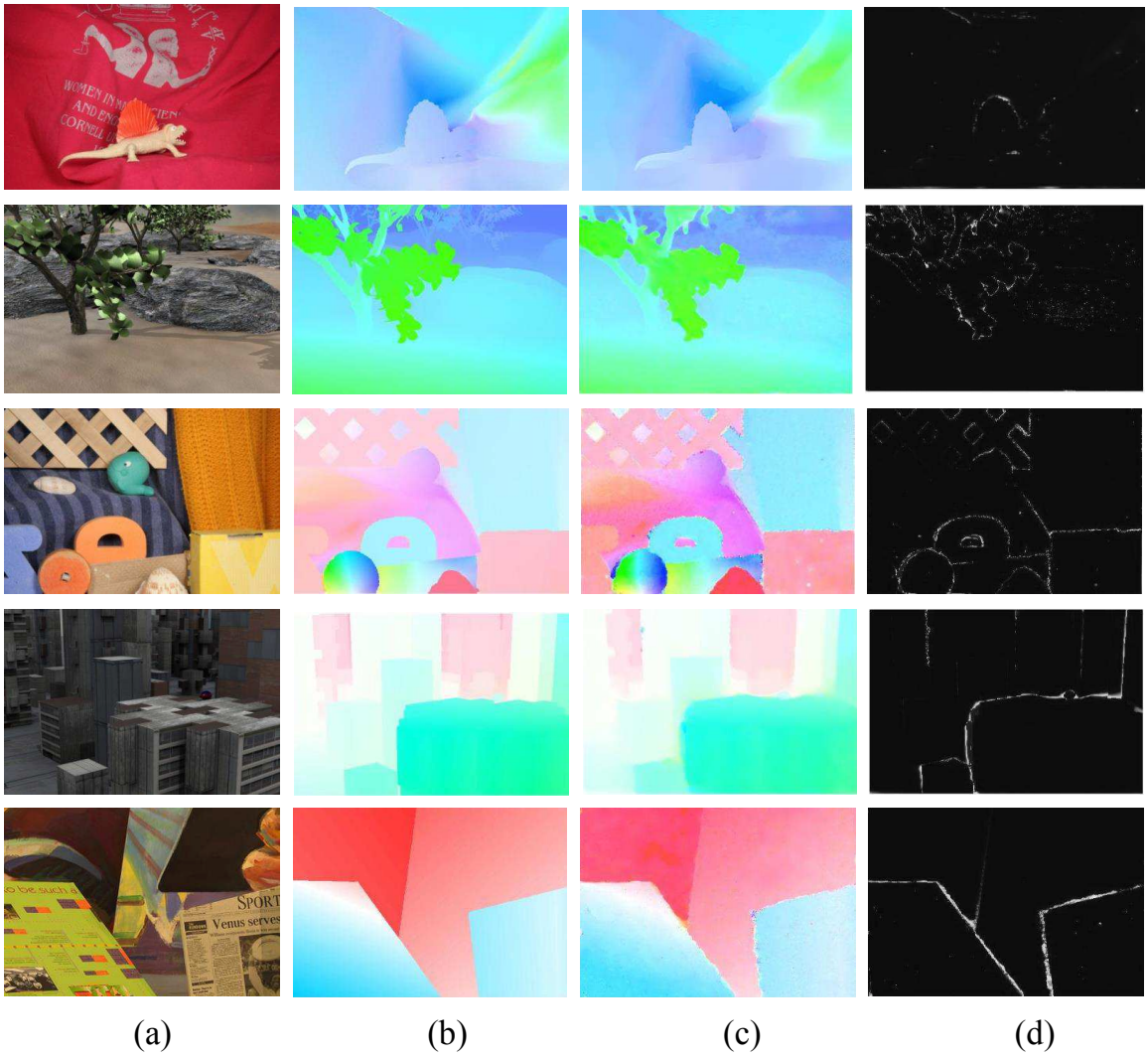


Figure 4.1: Optical flow estimation by the proposed method using the OFC. Each row corresponds to one video sequence. (a) Video frame, (b) the ground-truth flow and (c) the estimated optical flow. (d) The error between the ground-truth and the estimated optical flow. The color scheme used to indicate the direction and magnitude of the flow vectors is given in Figure 1.1 (e).

Table 4.2: MEPE results of the proposed regularizer with various data terms on Middlebury training dataset.

Data terms	**OFC**	**GCA** [78]	**GDIM** [98]	**NDC** [132]	**N-OFC** [30]
Dimetrodon	0.15	0.17	0.21	0.16	**0.12**
Grove2	0.12	**0.14**	0.18	0.21	0.15
Grove3	0.41	0.55	**0.43**	0.52	0.48
Hydrangea	**0.16**	0.18	0.21	0.22	0.25
Rubberwhale	0.16	**0.10**	**0.10**	0.18	0.15
Urban2	**0.41**	0.48	0.43	0.54	0.58
Urban3	0.65	**0.45**	0.51	0.62	0.49
Venus	0.28	**0.26**	0.32	0.34	0.29

sequences are used for optical flow estimation. The rest of the algorithm settings remain the same.

Figure 4.2 shows optical flow estimation results on real world sequences: *Flower garden*, *Dumptruck*, *Walking* and *Camera motion*. In *Flower garden* sequence, a moving camera captures the static scene. *Dumptruck* and *Walking* are chosen from Middlebury test dataset. The camera is static in the former, whereas both camera and person are moving in the latter. We do not have ground-truth for aforementioned real sequences because obtaining ground-truth motion data for real video sequences is a challenging task. However, *Camera motion* video sequence has been selected because its ground-truth optical flow is available. This sequence has been taken from annotated sequence database [165]. In Camera motion sequence, a moving camera captures the video of a complex multi-object traffic scene. These results are visually plausible because ti can be observed that motion boundaries are preserved in all four sequences.

4.4.3 Optical Flow Estimation from Reduced Measurements

In these experiments, the performance of the proposed method given in (4.16) is tested under reduced measurements of intensity derivatives. Three different schemes are used for the selection of intensity derivatives:

- selection of derivatives randomly;

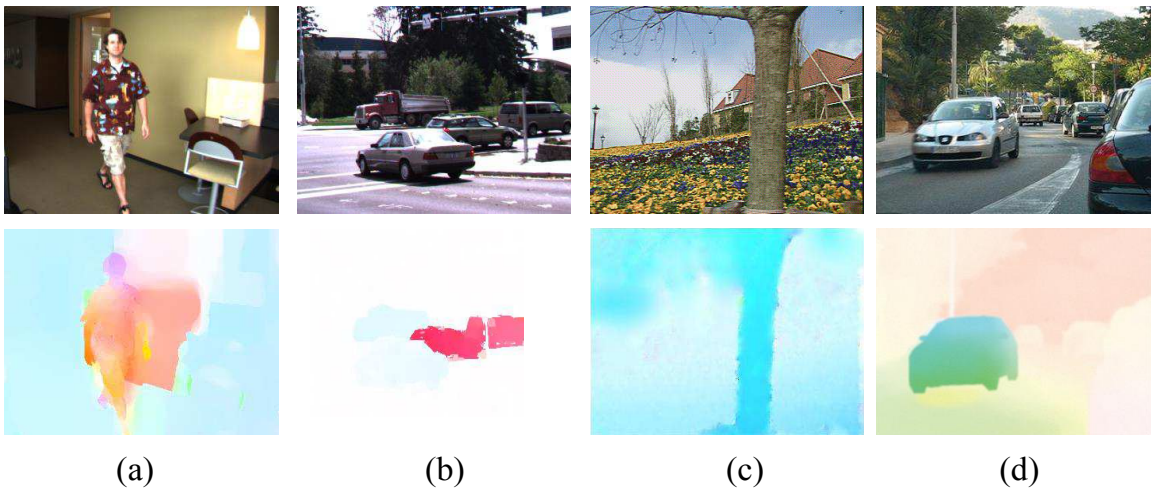


Figure 4.2: Optical flow estimation on real sequences (a) *Walking*, (b) *Dump truck*, (c) *Flower garden* and (d) *Camera motion*. Video frames are shown in the top row whereas estimated optical flows are in the bottom.

- selection of those derivatives which have significant magnitude;
- selection of a combination of random and significant derivatives.

Let m be the number of measurements in an n pixel video frame, we define the measurement ratio as m/n. The performance is evaluated for different values of the measurement ratio. Note that if $m/n = 1$, then full measurements are used, i.e., there is no reduction. These measurements are gradually increased in each type of selection, and the MEPE is plotted against each set of measurement ratio. In the case of random and combined selection schemes, the algorithm is run five times for each set of measurements, and the average of results is taken at the end. The regularization parameter λ has been set to 0.01 for these experiments.

We have also compared the number of measurements required to estimate optical flow with the sparsity (percentage of nonzero elements) of partial flow derivatives. To this end, the magnitude maps of the ground-truth partial flow derivatives and the ground-truth flow gradient are calculated. Otsu's global thresholding method is used to convert these maps into binary maps, then the number of nonzero in each result is counted to compute the sparsities of these maps. We have used Middlebury's training sequences for

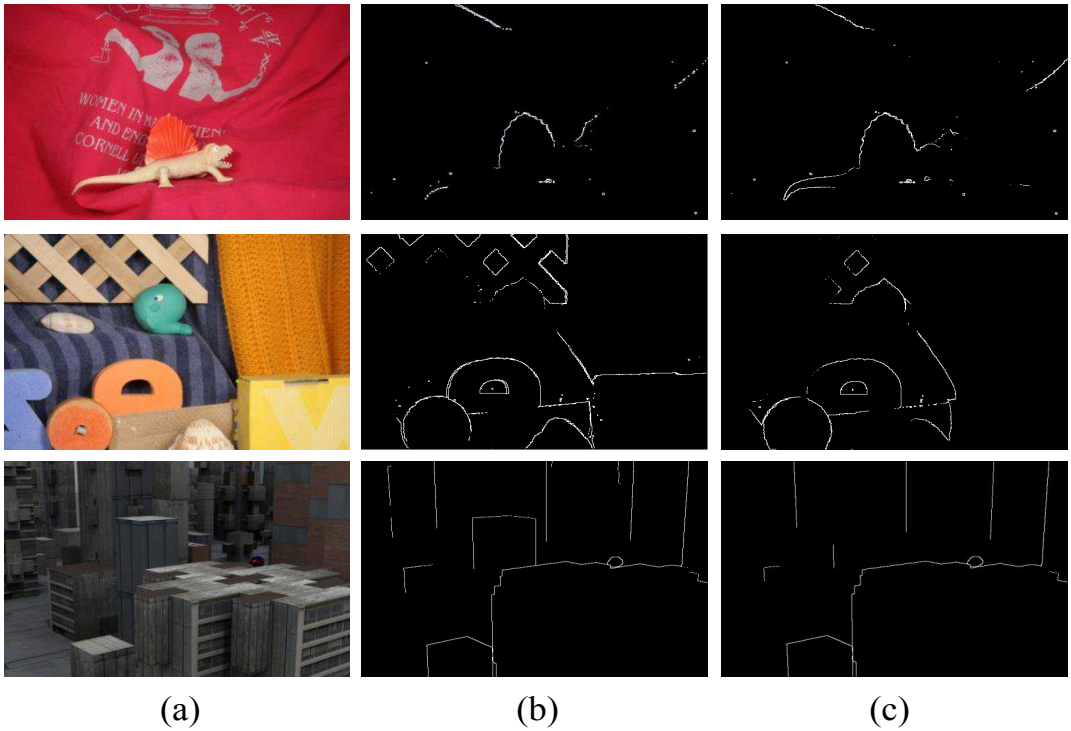


Figure 4.3: The sparsity of the ground-truth flow. (a) Video frame, the gradient magnitude map of (b) the horizontal $\mathbf{v}_{x\mathrm{GT}}$ and (c) vertical $\mathbf{v}_{y\mathrm{GT}}$ ground-truth flows.

these experiments because their ground-truth is available.

Figure 4.3 (a), (b) and (c) show the gradient flow magnitude maps of the horizontal and vertical ground-truth flows computed over 3 video sequences. These edge maps form motion boundaries, which are shown as white color in Figure 4.3(b) and (c). It can be seen that motion boundaries are formed from a small percentage of total number of pixels. Figure 4.4 clearly shows that only less than 5% of pixels form nonzero partial flow derivative magnitude and gradient flow magnitude maps in these video frames. Note that magnitude maps of partial flow derivatives $|\boldsymbol{\nabla}_x \mathbf{v}_{\mathrm{GT}}|$, $|\boldsymbol{\nabla}_y \mathbf{v}_{\mathrm{GT}}|$, $|\boldsymbol{\nabla}_{xy} \mathbf{v}_{\mathrm{GT}}|$ and $|\boldsymbol{\nabla}_{yx} \mathbf{v}_{\mathrm{GT}}|$ are sparser than the gradient flow magnitude map $|\boldsymbol{\nabla} \mathbf{v}_{\mathrm{GT}}|$.

Figure 4.5 presents the MEPE as a function of the measurement ratio for three different types of selections of intensity derivatives. In case of significant selection, and a measurement ratio of 0.4 or less, an increase in the error can be observed for most of video sequences. When a scene contains moving objects with strong edges and others

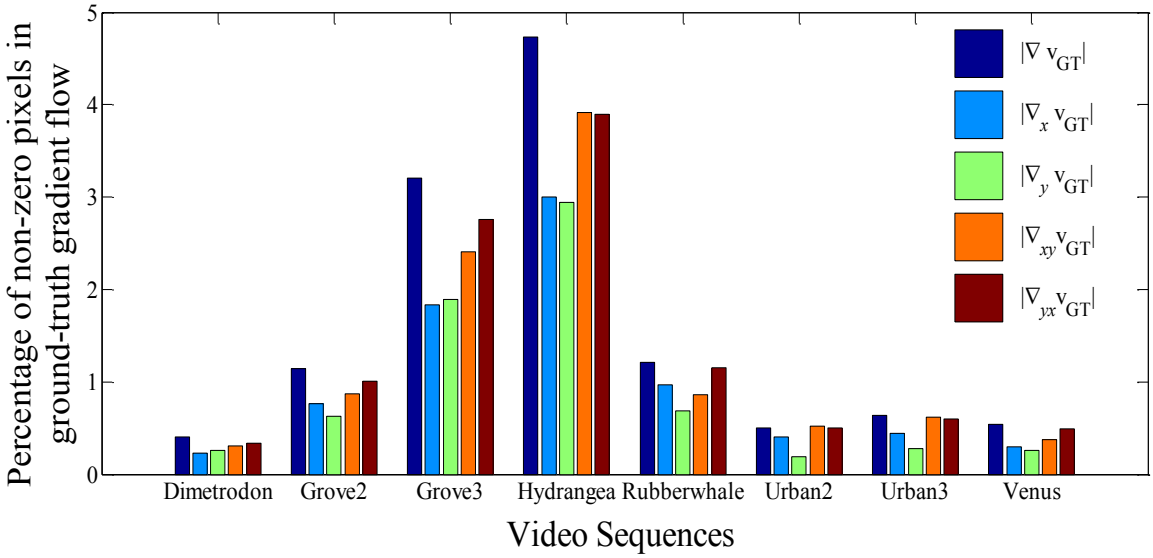


Figure 4.4: The percentage of nonzero pixels in derivative maps of the ground-truth flow. These eight video sequences are taken from Middlebury's training video sequences dataset.

with weak edges, then sensing only significant intensity derivatives fails to detect moving objects with weak edges. Consequently, it results in a higher value of the MEPE.

The error remains almost constant when the proposed method uses random selection of intensity derivatives up to $m/n = 0.1$, except for *Grove3* and *Rubberwhale* sequences. *Grove3* sequence has motion of small objects, and a measurement ratio of 0.1 corresponding to randomly selected derivatives misses details of small moving objects. Therefore, the error rises in this sequence, see Figure 4.5(b). *Rubberwhale* sequence has nonzero partial flow derivatives comparable to *Grove2*, *Urban2*, *Urban3* and *Venus* sequences, but it contains motion of several independently moving objects. Thus, unlike these sequences, the error rises in *Rubberwhale* for a measurement ratio of 0.1 as shown in Figure 4.5(d). *Hydrangea* sequence has the most number of nonzero partial flow derivatives in the ground-truth flow compared to other sequences as shown in Figure 4.4, but this sequence has only two distinct motions of Hydrangea bush and the background. Thus, optical flow is recovered correctly for a measurement ratio of 0.1. Unlike significant and random sensing, the combined sensing correctly estimates optical flow in all video sequences even when $m/n = 0.1$.

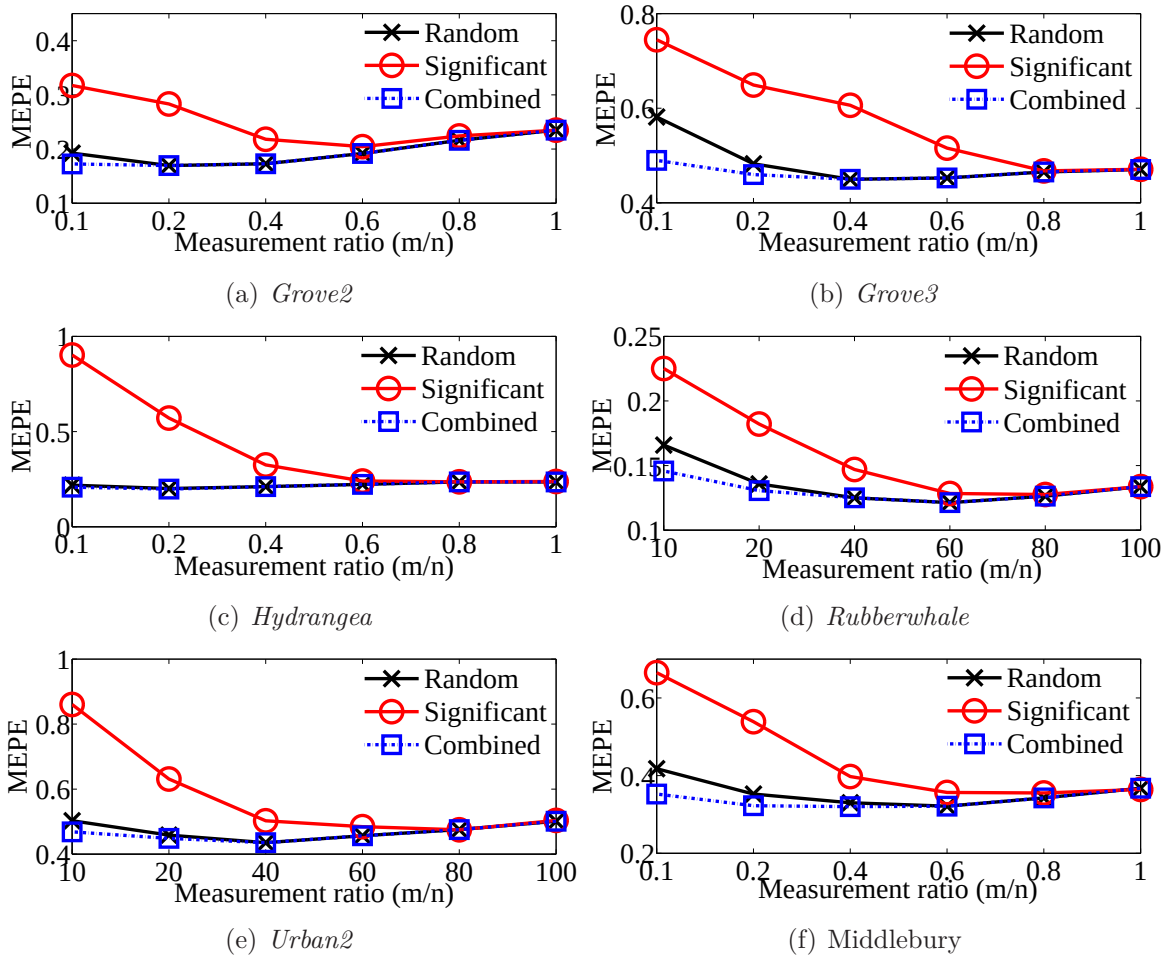


Figure 4.5: The MEPE as a function of the measurement ratio for three different selection schemes. Results are reported for (a) *Grove2*, (b) *Grove3*, (c) *Hydrangea*, (d) *Rubberwhale* and (e) *Urban2* sequences. An average MEPE over all videos of (f) Middlebury training database is also shown.

4.4.4 Comparison with Existing Methods

In these experiments, the performance of the proposed method is compared against state-of-the-art optical flow methods. Anisotropic TV regularization based method **TV-L1-improved** [2], image structure adaptive regularization based method **Adaptive** [3], structure adaptive TV and neighborhood dependent constraint method **Ad-TV-NDC**, and nonlocal TV regularization based method **Classic + NL** [4] have been selected for the comparison. Moreover, to test the effect of adaptive regularization, we have implemented

Table 4.3: Performance comparison of the proposed method with existing methods on Middlebury training dataset. The structure adaptive regularization is used with the proposed method, referred to as **Adaptive HVD**, and added to the comparison.

Methods	**TV-L1 improved** [2]	**Adaptive** [3]	**Classic + NL** [4]	**Ad-TV-NDC** [132]	**Proposed HVD**	**Adaptive HVD**
Dimetrodon	0.18	0.16	**0.13**	0.17	0.15	0.14
Grove2	0.19	0.16	0.16	0.17	0.15	**0.13**
Grove3	0.62	0.51	0.48	0.58	0.41	**0.39**
Hydrangea	0.25	0.21	**0.15**	0.25	0.16	0.17
Rubberwhale	0.21	0.18	0.15	0.16	0.12	**0.11**
Urban2	0.57	0.49	0.39	0.46	0.41	**0.37**
Urban3	0.77	0.64	**0.52**	0.68	0.56	0.59
Venus	0.38	0.33	0.29	0.32	0.28	**0.26**
Average	0.39	0.34	0.28	0.33	0.25	**0.24**

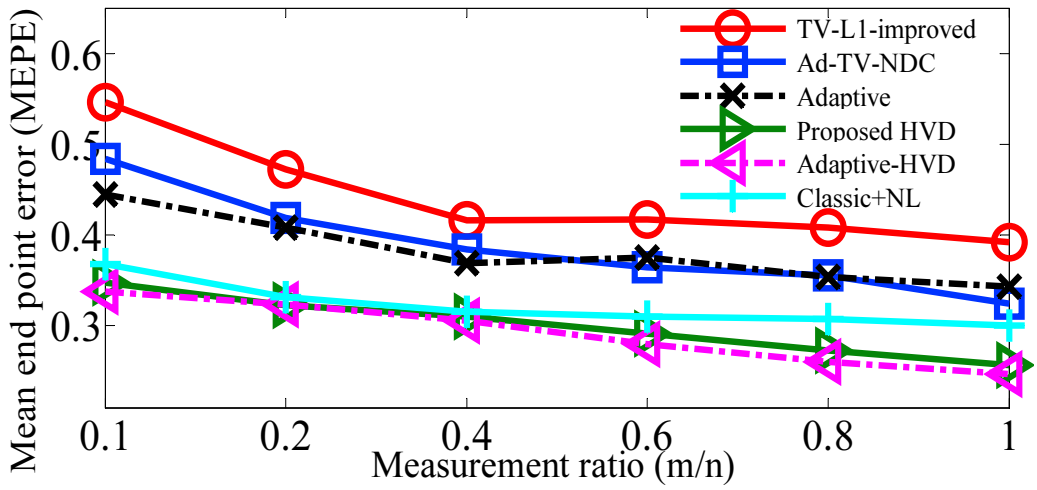
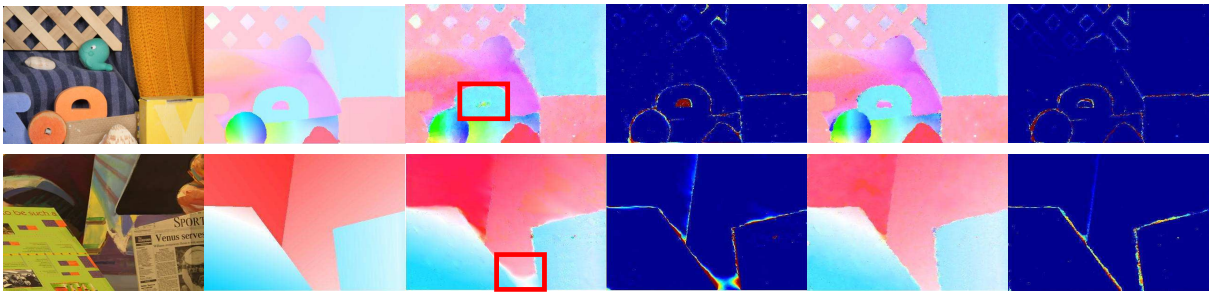


Figure 4.6: The MEPE as a function of the measurement ratio, computed on (a) *Dimetrodon*, (b) *Grove3* and (c) *Venus*. An average MEPE over all videos of (d) Middlebury training database is also computed. The proposed method using combined sensing is compared against **TV-L1-improved** [2], **Adaptive** [3], **Ad-TV-NDC** and **Classic + NL** [4]. The structure adaptive regularization is also used with the proposed method, referred to as **Adaptive-HVD**, and added to the comparison.

structure adaptive regularization in the proposed method. The same adaptive weighting scheme is used as proposed in [3]. It is referred to as **Adaptive HVD** in these experiments. The comparison is also conducted under reduced measurements. Since a combination of random and significant intensity derivative sensing performs better, this comparison is carried out using only the combined sensing scheme. In the combined sensing, we have used $0.05n$ significant derivatives, whereas the rest of the measurements are taken randomly.

Table 4.3 shows a performance comparison of the proposed method with aforementioned methods on Middlebury training dataset. It can be observed that **TV-L1-improved**, **Adaptive** and **Ad-TV-NDC** have higher MEPE than other three methods. For most of videos, the proposed method using the adaptive regularization **Adaptive HVD** outperforms other methods. **Classic + NL** ranks second in the comparison. However, the performance of **Perposed HVD** is comparable to **Classic + NL**.

Figure 4.6 presents the MEPE as a function of the measurement ratio for different



(a) Video frame (b) Ground-truth (c) Estimated flow (d) Estimation error (e) Estimated flow (f) Estimation error

Figure 4.7: The estimated optical flow and the estimation error for $m/n = 0.2$ by using a combined sensing scheme. Each row corresponds to one video sequence. (a) Video frame, (b) the ground-truth optical flow, the estimated optical flow and the resulting flow estimation error for the methods (c)-(d) **Adaptive** and (e)-(f) **proposed HVD**, respectively.

methods on Middlebury training dataset. The measurement ratio is gradually increased in these methods by using a combined sensing scheme. The error remains almost constant in case of **Proposed HVD** and **Adaptive-HVD** up to $m/n = 0.1$. It can be observed that there is not much improvement in terms of the error performance when adaptive regularization is used along with the HVD regularizer in the proposed method. This implies that the HVD regularizer, due to its anisotropic nature, can be used without adaptive weighting to preserve sharp motion boundaries. **Classic + NL** method, in general, gives similar performance as the proposed method. This method constrains optical flow by imposing multiple constraints in a neighborhood around each flow vector, and also uses occlusion detection techniques.

Figure 4.7 shows the estimated optical flow and the estimation error obtained by the proposed method and Adaptive for a measurement ratio of 0.2. The estimated optical flow by the method Adaptive in Figure 4.7(c) indicates that motion details in the highlighted portion of Rubberwhale video sequence have been lost. Furthermore, excessive blurring can be seen in the lower middle part of Venus sequence by the same method. On the contrary, the proposed method does not experience these problems, as shown in Figure 4.7 (b) and (e).

We have also compared the run time of the proposed method with **Adaptive** and

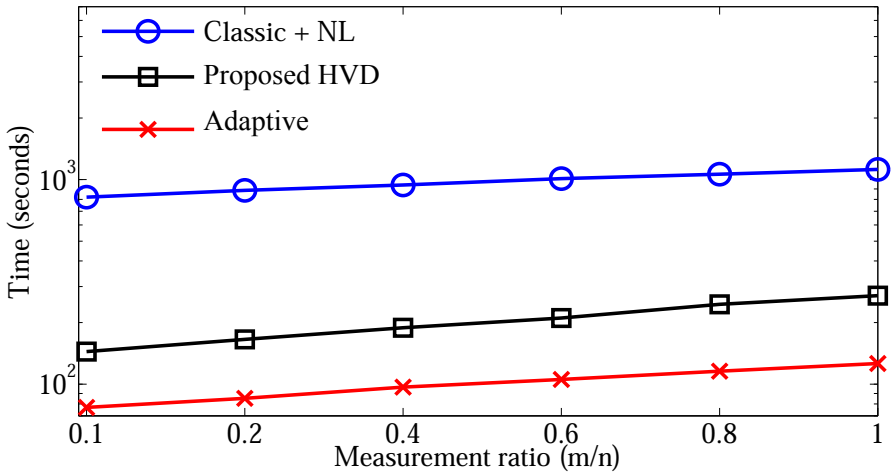


Figure 4.8: Run time as a function of the measurement ratio for **Adaptive**, **Proposed HVD** and **Classic + NL** methods.

Classic + NL methods against the number of measurements. Figure 4.8 shows the time taken to compute optical flow by these methods. Our experiments show that by using 20% of the measurements in Adaptive, we reduce the computation time by up to 30%. Moreover, about 20% time gain is obtained for same number of measurements in the proposed method. Though the performance of **Classic + NL** is comparable to the proposed method, we have found that this method is much slower compared to **Proposed HVD** and **Adaptive**. **Classic + NL** takes the most time because minimization of nonlocal terms lead to a higher computational cost in this method.

4.5 Discussion and Conclusions

This chapter utilizes the proposed HVD regularizer for the estimation of dense optical flow. The proposed regularizer exploits the sparsity of the gradient field of motion vectors in the spatial domain. It imposes the continuity of partial flow derivatives in a small neighborhood around each flow vector, and minimizes them separately.

We have used five different data terms with the proposed method. First, optical flow constraint is used to construct a data term. Second, a neighborhood dependent constraint is used for the design of data term. However, these two data terms can not handle

brightness change in digital videos. To estimate optical flow under changing brightness conditions of underlying videos, three data terms are used. Normalized OFC, gradient constancy assumption and a generalized dynamic image model based data terms are embedded with the proposed regularizer to handle brightness change.

Experiments have been conducted to estimate optical flow from synthetic as well as real video sequences. Results presented in this chapter show that the proposed method is able to reduce undesirable staircase artefacts as generated by TV based regularization of optical flow. The proposed method also preserves sharp horizontal, vertical as well as diagonal motion boundaries. Though computationally less expensive, its performance is comparable to the nonlocal total variation.

To estimate optical flow from highly incomplete measurements, three different types of sensing schemes are used for the selection of spatiotemporal intensity derivatives: significant, random and the combined sensing. Our experiments show that the combined sensing of spatiotemporal intensity derivatives performs better than random only or significant only sensing. By using a combined sensing scheme and the HVD regularizer, it has been shown that optical flow can be estimated by using only 10% of total measurements without too much sacrificing the accuracy of the estimation.

Chapter 5

Joint Optical Flow Estimation and Sparse Occlusion Detection under Varying Illumination

Chapter contents

5.1 Introduction

In any video sequence, it is most likely that moving objects will obstruct the view of other objects. The phenomenon of occlusion causes the blockage of some scene points, and reveals some new portions of the scene across video frames. Occlusions play an important role in the visual perception of the three-dimensional shape of objects and the scene structure [104]. Occlusions are usually observed at moving object boundaries and discontinuities in the scene depth. The knowledge of these occlusions can be effectively used in object segmentation as well as in video coding techniques to improve their efficiency. Low-level image processing techniques and high-level vision tasks also get benefit from occlusion detection methods [6].

The matching of the image pixels from one video frame to another is possible only when objects in the scene remain visible over time. Optical flow, which is used to match brightness patterns across video frames, is not defined in occluded regions. Moreover, the matching of the brightness patterns may not be possible in the presence of varying illumination conditions. Consequently, optical flow estimation methods can wrongly detect occlusion in those visible video regions which exhibit the violation of the brightness constancy.

In this chapter, the problem of occlusion detection using optical flow is solved under changing illumination of the scene. Since occluded video regions form a small portion in any video frame, the method proposed in this chapter imposes a sparsity prior on these regions for joint optical flow estimation and occlusion detection. The problem of sparse occlusion detection with optical flow has been addressed in [5], which assumes Lambertian surface and constant illumination of the scene. The OFC is then used under these assumptions for the joint estimation of optical flow and occlusion detection. Our work is close in spirit to this approach; however, unlike the method presented in [5], the proposed method can work under brightness change. The proposed method uses a generalized dynamic image model (GDIM) [98] in place of the OFC to handle brightness change. The GDIM captures the violation of brightness constancy assumption, caused by diffuse

shading and specular flows, in its parameters. These parameters are also estimated along with joint optical flow estimation and occlusion detection in this chapter. An extension of the proposed method has also been presented which does not require the estimation of brightness change parameters. Therefore, this modified method is computationally less expensive than the original proposed method at the cost of a small decrease in optical flow estimation and occlusion detection performances.

The rest of the chapter is organized as follows. Section 5.2 describes optical flow methods to detect occluded video regions for Lambertian reflectance under homogeneous illumination. Section 5.3 proposes the method to jointly estimate optical flow and detect sparse occlusion under varying illumination. It also presents a modification to the proposed method, which does not require the estimation of brightness change parameters under certain conditions. Section 5.4 gives algorithmic details of the proposed method. Finally, Section 5.5 presents the experimental setup and results to evaluate the performance of the proposed method.

5.2 Occlusion Detection using Optical Flow under Homogeneous Illumination

Optical flow is undefined in occluded video regions; therefore, the data term can be used to detect occlusion where it is violated as a result of the appearance or the disappearance of scene points. This section describes the methods which use the residual of the data term for occlusion detection. Since occluded regions between two consecutive video frames usually form a small portion in any video frame, a sparsity constraint can be applied to these regions. This section also presents sparse occlusion detection technique using optical flow.

5.2.1 The Residual of the Data Term

Since the data term is used to match the scene points across the image sequences, its residual can give a clue about occluded regions when the matching of the points is not

possible. A pixel is taken as occluded when the residual of the data term exceeds a threshold. The residual of the OFC, for example, can be used for occlusion detection, which is given as

$$\rho(\vec{v}) = \begin{cases} 1 & \text{if } \Big(\big(I_1(x,y,t) - I_2(x-v_x, y-v_y, t-dt)\big)^2 > \epsilon_I, \\ 0 & \text{otherwise,} \end{cases} \tag{5.1}$$

where the threshold ϵ_I decides whether the pixel is occluded or not, $\rho = 1$ indicates the occlusion, whereas $\rho = 0$ means that the pixel is visible across video frames. Equation (5.1) is an example of hard thresholding. To obtain a continuous differentiable function $\rho(\vec{v})$, a soft thresholding can be used in place of hard thresholding. In [112], for example, the following soft thresholding is used

$$\rho(\vec{v}) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\Big(\big(I_1(x,y,t) - I_2(x-v_x, y-v_y, t-dt)\big)^2 - \epsilon_I\Big), \tag{5.2}$$

which is a differentiable approximation of the Heaviside function. The data term can also be a modified form of the OFC. In [113], for example, the video motion is represented as particles. Brightness patterns are matched by using a data term, which resembles the OFC. This data term is combined with the divergence of optical flow to detect occlusion.

Occlusions often occur at image discontinuities. Therefore, some methods use image driven anisotropic diffusion to detect occluded image regions by exploiting the motion symmetry [100, 110, 111]. This type of methods compute forward as well as backward flow, and use a consistency function between both flows. This consistency function depends upon the residual of the data term; video regions with low consistency values are taken as occluded regions.

The cross-checking of the motion symmetry can be effective in occlusion detection, but these methods compute optical flow bi-directionally. Another class of methods are based on an observation that pixels in the reference video frame should uniquely correspond to other pixels in the target frame [107, 108]. The mapping of multiple pixels in the target frame to the same pixel in the reference frame can be used as an indication of occluded pixels. This mapping uniqueness criteria also uses the residual of the data term, and has been used to label occluded pixels in [82].

5.2.2 Sparse Occlusion Detection

This section reviews sparse occlusion detection using optical flow estimation method presented in [5]. This method serves as a basis for our proposed method. It considers occluded regions to be sparse, and uses convex optimization techniques to detect them. The ℓ_1 norm of the residual of the OFC is minimized in a variational framework to detect sparse occlusion regions. The method assumes constant illumination of the scene; therefore, it works on the assumption that the change in the image brightness is only due to discrete displacements of pixels from one video frame to another. This brightness constancy assumption is used as a matching term in a variational energy minimization framework.

Let us define a time-varying grayscale image $I(x, y, t)$ on a domain $\mathfrak{D}$, where $(x, y) \in \mathfrak{D}$ are spatial coordinates and t is the time variable. Moreover, we define $\mathfrak{R}$ and $\mathfrak{R}'$ to be co-visible and occluded video regions such that $\mathfrak{R} \cup \mathfrak{R}' = \mathfrak{D}$ and $\mathfrak{R} \cap \mathfrak{R}' = \emptyset$. Under the assumption of Lambertian reflectance and constant illumination, the OFC for two consecutive video frames can be expressed as

$$I(x, y, t) = \begin{cases} o(x, y, t), & (x, y) \in \mathfrak{R}', \\ I(x - v_x dt, y - v_y dt, t - dt) + e(x, y, t), & (x, y) \in \mathfrak{R}, \end{cases} \tag{5.3}$$

where $o(x, y, t)$ is an occlusion term, $e(x, y, t)$ represents the error in the measurement process, dt is the temporal sampling interval, v_x and v_y are, respectively, the horizontal and vertical flow components between video frames at time t and $t - dt$. The occlusion term $o(x, y, t)$ is generally unrelated to $I(x - v_x dt, y - v_y dt, t - dt)$, and the image can take on any value inside $\mathfrak{R}'$. Therefore, $o(x, y, t)$ should not be matched with the initial image brightness $I(x, y, t)$. The term $e(x, y, t)$ accumulates a large number of independent phenomena such as un-modeled illumination changes in the scene, measurement error, quantization error and sensor noise, to name a few. Thus, this term is assumed here as an independent and identically distributed random variable $e(x, y, t) \sim \mathcal{N}(0, \sigma)$ with the standard deviation σ. The residual on the entire image domain $\mathfrak{D}$ can now be defined as

$$r(x, y, t; dt) := I(x, y, t) - I(x - v_x dt, y - v_y dt, t - dt). \tag{5.4}$$

Using this definition in (5.3), residuals in occluded video regions $r_1(x, y, t; dt)$ and co-visible regions $r_2(x, y, t; dt)$ can be given as

$$\begin{pmatrix} r_1(x, y, t; dt) \\ r_2(x, y, t; dt) \end{pmatrix} = \begin{cases} o(x, y, t) - I(x - v_x dt, y - v_y dt, t - dt), & (x, y) \in \mathfrak{R}', \\ e(x, y, t), & (x, y) \in \mathfrak{R}. \end{cases} \tag{5.5}$$

Note that occluded region $r_1(x, y, t; dt)$ is undefined in $\mathfrak{R}$, and co-visible region $r_2(x, y, t; dt)$ is undefined in $\mathfrak{R}'$. Therefore, they can take on any value where they are undefined, including zero, which is assumed in this model. The OFC can now be written as the sum of two residual terms in co-visible and occluded video regions:

$$I(x, y, t) - I(x - v_x dt, y - v_y dt, t - dt) = r_1(x, y, t; dt) + r_2(x, y, t; dt). \tag{5.6}$$

The term $I(x - v_x dt, y - v_y dt, t - dt)$ in (5.6) is a nonlinear term in v_x and v_y. It can be linearized when dt and optical flow between t and $t - dt$ are sufficiently small. The linearized OFC can be written as

$$(\nabla I(x, y, t))^{\mathrm{T}} \vec{v}(x, y, t) + I_t(x, y, t) = r_1(x, y, t; dt) + r_2(x, y, t; dt), \tag{5.7}$$

where $\nabla = [\nabla_x \ \nabla_y]^{\mathrm{T}}$ is a gradient operator, $\nabla I(x, y, t) = [I_x \ I_y]^{\mathrm{T}}$ is the spatial image gradient, $I_t(x, y, t)$ is the temporal brightness derivative, and $\vec{v}(x, y, t) = [v_x \ v_y]^{\mathrm{T}}$ is a two-dimensional velocity vector. The method uses Equation (5.7) to construct a data term $E_{\text{data}}(\vec{v}(x, y, t), r_1(x, y, t; dt))$ for the matching of the brightness of moving patterns across video frames. In any image sequence, $r_1(x, y, t; dt)$ is generally sparse and $r_2(x, y, t; dt)$ is dense; therefore, the inference criteria for $\vec{v}(x, y, t)$ and $r_1(x, y, t; dt)$, used in [5], can be given as

$$E_{\text{data}}(\vec{v}, r_1) = \alpha ||r_1||_0 + ||r_2||_2 \qquad \text{subject to} \qquad (\nabla I)^{\mathrm{T}} \vec{v} + I_t \ = \ r_1 + r_2, \tag{5.8}$$

where α is a positive weighting factor for the occlusion term $r_1(x, y, t; dt)$, and $||.||_0$ is the so called ℓ_0 "norm" of $r_1(x, y, t; dt)$. Note that we have omitted the spatiotemporal dependence of terms for the sake of brevity. Equation (5.8) can also be written as

$$E_{\text{data}}(\vec{v}, r_1) = \alpha ||r_1||_0 + ||(\nabla I)^{\mathrm{T}} \vec{v} + I_t - r_1||_2. \tag{5.9}$$

Brightness patterns across video frames are matched by the second term in (5.9). This term imposes a pointwise similarity on the unknown $\vec{v}$; it can only estimate the normal velocity component. Thus, it can not be used alone to estimate optical flow. This is known as the *aperture problem* in optical flow literature. To cope with the aperture problem, some extra information needs to be incorporated. Variational methods usually impose a smoothness constraint over optical flow to regularize the problem. Nonlinear edge preserving smoothness, for example, the weighted total variation $||\nabla\vec{v}||_{\mathrm{TV}}$ can be used to regularize optical flow as

$$E_{\mathrm{reg}}(\vec{v}) = \lambda_1||\nabla\vec{v}||_{\mathrm{TV}} = \lambda_1 \sum_{i,j} |\nabla v_{x(i,j)}| + |\nabla v_{y(i,j)}|, \tag{5.10}$$

where λ_1 is a regularization parameter and $|\nabla v_{x,y}| = \sqrt{(\nabla_x v_{x,y})^2 + (\nabla_y v_{x,y})^2}$. To control the weight of the regularizer, the regularization parameter λ_1 can be made adaptive to the structure of the image:

$$\lambda_1 = \lambda \exp(-\gamma|| \sqrt{(\nabla_x I)^2 + (\nabla_y I)^2}||_2) + \kappa, \tag{5.11}$$

where λ is a constant, γ controls the variations in the weight of the regularizer according to the image structure, and a small constant κ is used to avoid the weight of the regularizer to become zero in flat regions where there is a little intensity change. The regularization parameter λ_1 gives less weight to the regularizer at image boundaries which are likely to be motion boundaries; thus, it is expected to reduce the undesirable blurring of motion boundaries. The joint optical flow estimation with sparse occlusion detection can now be formulated as the minimization of the data energy $E_{\mathrm{data}}(\vec{v}, r_1)$ and the regularization energy $E_{\mathrm{reg}}(\vec{v})$:

$$E(\vec{v}, r_1) = \min_{\vec{v}, r_1} \alpha||r_1||_0 + ||(\nabla I)^{\mathrm{T}}\vec{v} - r_1 + I_t||_2 + \lambda_1||\nabla\vec{v}||_{\mathrm{TV}}. \tag{5.12}$$

The domain $\mathfrak{D}$ is discretized for the digital video sequences; thus, (5.12) can be written in the matrix form. For an n-dimensional lexicographically vectorized video frame, let $\mathbf{v} \in \mathbb{R}^{2n\times 1}$ and $\mathbf{r}_1 \in \mathbb{R}^{n\times 1}$ represent the columnwise stacked $2n$-dimensional optical flow

and n-dimensional occlusion term, respectively. Moreover, let the vector $\mathbf{y}_1$ corresponds to the stacked temporal brightness derivative I_t in a similar fashion. The matrix form of (5.12) is now given as

$$E(\mathbf{v}, \mathbf{r}_1) = \min_{\mathbf{v},\mathbf{r}_1} \alpha ||\mathbf{r}_1||_0 + ||\Big[\text{diag}(\mathbf{I}_x)\ \text{diag}(\mathbf{I}_y)\Big]\mathbf{v} - \mathbf{r}_1 + \mathbf{y}_1||_2 + \lambda_1 ||\boldsymbol{\nabla}\mathbf{v}||_{\text{TV}}, \tag{5.13}$$

where $\mathbf{I}_x$ and $\mathbf{I}_y$ are, respectively, columnwise stacked horizontal and vertical intensity derivatives, $\text{diag}(\mathbf{I}_x)$ and $\text{diag}(\mathbf{I}_y)$ are corresponding diagonal matrices, and $\boldsymbol{\nabla} = [\boldsymbol{\nabla}_x\ \boldsymbol{\nabla}_y]^{\text{T}}$ is the matrix to calculate the horizontal and vertical derivatives of the vectorized optical flow $\mathbf{v}$. Let us define the matrix A to be a concatenation of three $n \times n$ matrices as

$$A = [\text{diag}(\mathbf{I}_x)\ \text{diag}(\mathbf{I}_y)\ \ -\mathcal{I}]^{\text{T}},$$

where $\mathcal{I}$ is the identity matrix. Equation (5.13) can be written as

$$E(\mathbf{v}, \mathbf{r}_1) = \min_{\mathbf{v},\mathbf{r}_1} \alpha ||\mathbf{r}_1||_0 + ||A\mathbf{u} + \mathbf{y}_1||_2 + \lambda_1 ||\boldsymbol{\nabla}\mathbf{v}||_{\text{TV}}, \tag{5.14}$$

where $\mathbf{u} = [\mathbf{v}\ \mathbf{r}_1]^{\text{T}}$. The problem in (5.14) is NP-hard as it involves the so called "ℓ_0 norm" minimization of the occlusion term $\mathbf{r}_1$. Alternatively, under certain conditions, the ℓ_0 is replaced by a convex ℓ_1 norm [166]. Since the ℓ_1 norm relaxation may undesirably penalize strong (large residual of the OFC) occluded regions, it is more suitable to use a weighted ℓ_1 norm [167]. Optical flow and sparse occlusion can now be jointly estimated by solving the following energy functional:

$$\mathbf{u} = \arg\min_{\mathbf{v},\mathbf{r}_1} \alpha ||W\mathbf{r}_1||_1 + ||A\mathbf{u} + \mathbf{y}_1||_2 + \lambda_1 ||\boldsymbol{\nabla}\mathbf{v}||_{\text{TV}}, \tag{5.15}$$

where diagonal weight matrix W iteratively adapts weights for a better ℓ_0 approximation. It can be chosen as $W = \frac{1}{|\mathbf{r}_1|+\delta}$. The ℓ_1 norm reweighting with such a scheme significantly enhances the sparsity of the solution; therefore, it computes $\mathbf{r}_1$ close to an indicator function by less penalizing strong occluded pixels. Equation (5.15) is used to detect sparse occlusion using optical flow estimation under homogeneous illumination of the scene.

5.3 Sparse Occlusion Detection using Optical Flow under Varying Illumination

The OFC is violated and the mapping of the pixels across consecutive video frames may not be possible when various radiometric and geometric factors cause a change in the scene radiance and irradiance. The OFC is also violated in occluded regions as mentioned above; therefore, optical flow estimation methods may erroneously detect occlusion in the visible regions undergoing violation of the brightness constancy assumption. The violation in the brightness constancy assumption is caused by physical and radiometric factors: light source motion with respect to the scene, change in the moving surface orientation and shadows cast on moving objects are some of the examples. Figure 5.1 shows the effect of these violations in the OFC based optical flow estimation. The OFC results in incorrect matching of the scene points under varying illumination of the scene as shown in Figure 5.1(b) and (c).

This section presents the proposed method for sparse occlusion detection using optical flow estimation. First, to capture brightness constancy violations, a generalized dynamic image model (GDIM) is presented in Section 5.3.1. Since the matching term in the proposed method treats co-visible and occluded video regions separately, the GDIM is described for both regions. Second, the GDIM based brightness matching data term has been embedded into sparse occlusion detection method in Section 5.3.2. Third, a modification to the proposed method, which does not require the estimation of brightness change parameters of the GDIM, is presented in Section 5.3.3.

5.3.1 Generalized Dynamic Image Model (GDIM)

Light reflected from mirror like objects, and shadows cast over the scene points are two major sources of the violation of brightness constancy assumption. As a result, the estimated optical flow and the detected occlusion obtained by solving (5.15) can be severely biased and unreliable. To cope with the brightness constancy violation, the brightness change can be modeled as an affine function of the reference image brightness in the form

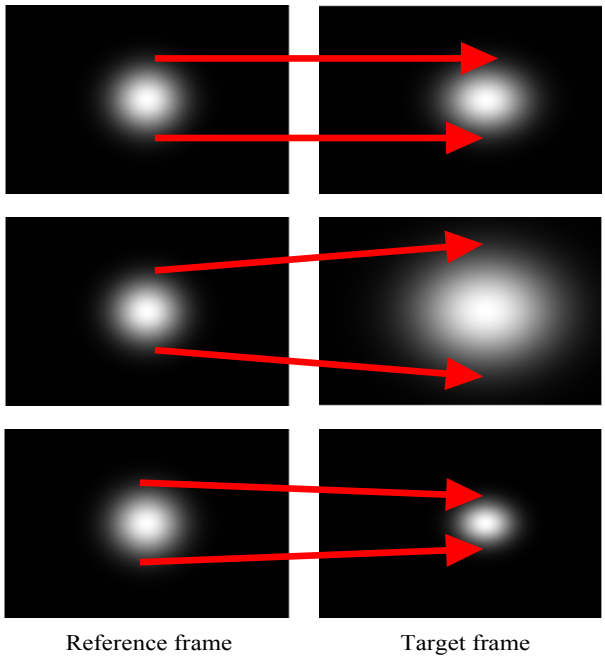


Figure 5.1: The OFC based optical flow estimation between the reference and target video frames. A bright pattern is translating from left to right in all three scenarios (top, middle and bottom). The correct matching of pixels is achieved (straight red arrows) in case of no brightness constancy violation (top) while wrong points are matched when there is an increase (middle) or decrease (bottom) in the brightness of moving pattern.

of a multiplier and an offset field [98]. This model can be described mathematically as

$$I(x,y,t) - I(x - v_x dt, y - v_y dt, t - dt) = d(x,y,t)I(x,y,t) + c(x,y,t), \tag{5.16}$$

where the multiplier field $d(x,y,t)$ permits changes in the image contrast and the offset field $c(x,y,t)$ allows for mean brightness changes from one video frame to another. The multiplier field is related to diffuse reflection; this type of reflection occurs when the incident light is reflected from rough surfaces. The reflection from mirror like surfaces, which is known as specular reflection, is modeled by the offset field $c(x,y,t)$ in this model. Please refer to Figure 5.2 for these two types of reflections.

Following occlusion modeling given in (5.3), the generalized dynamic image model (GDIM) for co-visible and occluded video regions can be given as

$$I(x,y,t) = \begin{cases} o(x,y,t), & (x,y) \in \Re', \\ I(x - v_x dt, y - v_y dt, t - dt) + d(x,y,t)I(x,y,t) & \\ +c(x,y,t) + e(x,y,t), & (x,y) \in \Re. \end{cases} \tag{5.17}$$

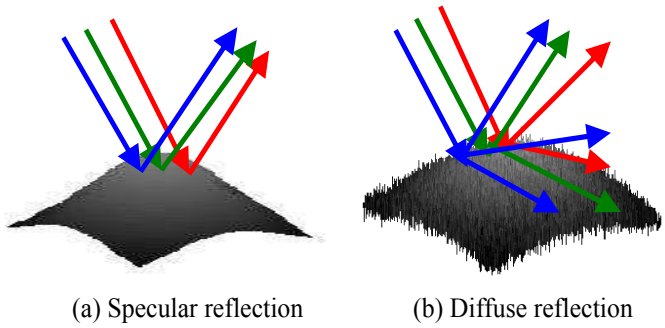


Figure 5.2: Light reflection from a glazing and a rough surface giving rise to (a) specular and (b) diffuse reflections, respectively.

The brightness of moving patterns, in co-visible video regions, satisfies the GDIM while $o(x, y, t)$ can take on any value because optical flow is undefined in occluded regions; therefore, occluded scene points are not matched across video frames.

5.3.2 Proposed method with Parameter Estimation

The proposed method uses the GDIM in place of the OFC to construct the data term. Since two brightness change parameters $d(x, y, t)$ and $c(x, y, t)$ appear in the GDIM, optical flow based sparse occlusion detection using the GDIM requires these parameters to be estimated. Therefore, to estimate the brightness change parameters of the GDIM, the proposed method constrains them properly. Similar to Equation (5.6), the GDIM for both co-visible $\mathfrak{R}$ and occluded $\mathfrak{R}'$ image regions can be written as

$$\begin{aligned} I(x, y, t) = \; & \Big(I(x - v_x dt, y - v_y dt, t - dt) + d(x, y, t) I(x, y, t) + c(x, y, t)\Big) \\ & + r_1(x, y, t; dt) + r_2(x, y, t; dt). \end{aligned} \tag{5.18}$$

Following the same procedure as in Section 5.2.2, the discretized data term employing the GDIM can be expressed as

$$\begin{aligned} E_{\text{data}}(\mathbf{v}, \mathbf{r}_1, \mathbf{d}, \mathbf{c}) = \alpha ||W \mathbf{r}_1||_1 + \\ || \begin{bmatrix} \text{diag}(\mathbf{I}_x) & \text{diag}(\mathbf{I}_y) & -\mathcal{I} & -\mathbf{I} & -\mathcal{I} \end{bmatrix} \begin{bmatrix} \mathbf{v}_x & \mathbf{v}_y & \mathbf{r}_1 & \mathbf{d} & \mathbf{c} \end{bmatrix}^{\text{T}} + \mathbf{y}_1 ||_2, \end{aligned} \tag{5.19}$$

where lexicographically vectorized $\mathbf{I}$, $\mathbf{v}_x$, $\mathbf{v}_y$, $\mathbf{d}$ and $\mathbf{c}$ are, respectively, the brightness of the reference video frame, the horizontal flow, the vertical flow, the multiplier field and

the offset field, and $\mathcal{I}$ is the identity matrix. Let us define

$$\mathbf{x} = \begin{bmatrix} \mathbf{v}_x & \mathbf{v}_y & \mathbf{r}_1 & \mathbf{d} & \mathbf{c} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} \mathbf{v} & \mathbf{r}_1 & \mathbf{d} & \mathbf{c} \end{bmatrix}^{\mathrm{T}}$$

and

$$B = \begin{bmatrix} \mathrm{diag}(\mathbf{I}_x) & \mathrm{diag}(\mathbf{I}_y) & -\mathcal{I} & -\mathbf{I} & -\mathcal{I} \end{bmatrix}^{\mathrm{T}}.$$

Equation (5.19) can be written as

$$E_{\mathrm{data}}(\mathbf{v}, \mathbf{r}_1) = \alpha ||W\mathbf{r}_1||_1 + ||B\mathbf{x} + \mathbf{y}_1||_2. \tag{5.20}$$

The parameter vector $\mathbf{x}$ is formed by the concatenation of brightness change parameters $\mathbf{d}$ and $\mathbf{c}$ with optical flow and the occlusion term. Since the estimation of $\mathbf{x}$ involves the estimation of brightness change parameters, we need to impose constraints on these parameters. When the digital videos are captured at sufficiently high spatial and temporal sampling rates, the natural causes of the brightness constancy violation do not appear suddenly in these videos. Therefore, it is quite safe to assume that the brightness constancy violation is not abrupt spatiotemporally [97, 168, 169]. The spatiotemporal variation of brightness change parameters is smooth in this case. To estimate $\mathbf{d}$ and $\mathbf{c}$, we impose a smoothness assumption over them. The term integrating the smoothness assumption over these parameters can be written as

$$E_{\mathrm{params}}(\mathbf{d}, \mathbf{c}) = \lambda_2 ||\boldsymbol{\nabla}\mathbf{d}||_2 + \lambda_3 ||\boldsymbol{\nabla}\mathbf{c}||_2, \tag{5.21}$$

where the weights λ_2 and λ_3 control the smoothing of these parameters. The overall energy to be minimized can now be given as

$$E(\mathbf{v}, \mathbf{r}_1, \mathbf{d}, \mathbf{c}) = E_{\mathrm{data}}(\mathbf{v}, \mathbf{r}_1) + E_{\mathrm{reg}}(\mathbf{v}) + E_{\mathrm{params}}(\mathbf{d}, \mathbf{c}), \tag{5.22}$$

and the parameter vector $\mathbf{x}$ is obtained by minimizing this energy as

$$\mathbf{x} = \arg \min_{\mathbf{v}, \mathbf{r}_1, \mathbf{d}, \mathbf{c}} \alpha ||W\mathbf{r}_1||_1 + ||B\mathbf{x} + \mathbf{y}_1||_2 + \lambda_1 [\mathrm{HVD}(\mathbf{v})] + \lambda_2 ||\boldsymbol{\nabla}\mathbf{d}||_2 + \lambda_3 ||\boldsymbol{\nabla}\mathbf{c}||_2. \tag{5.23}$$

Though data terms in (5.15) and (5.23) are different, the regularization term $E_{\mathrm{reg}}(\mathbf{v})$, however, remains the same. Note that (5.23) jointly estimates optical flow and detect occlusion

for two consecutive video frames because the generalized dynamic image model captures brightness constancy violations between these frames. This method can be extended to multiple video frames by modeling the brightness change nonlinearly, as presented in [97]. However, their approach also boils down to the linearized parameter estimation over time for complicated physical processes, for example, the brightness change caused by changing surface orientation and moving illumination envelopes, to name a few.

5.3.3 Proposed method without Parameter Estimation

It is sometimes useful to construct a data matching term that allows small variations in the value of the brightness. The brightness gradient can be one of the examples which permits small deviations from the brightness constancy assumption, and is relatively less sensitive to the brightness changes as compared to the brightness itself. Let us apply the gradient operator on both sides of (5.16); the gradient of the GDIM for co-visible and occluded video regions can now be given as

$$\nabla I(x,y,t) = \begin{cases} o(x,y,t), & (x,y) \in \mathfrak{R}', \\ \nabla\Big(I(x-v_x dt, y-v_y dt, t-dt) + d(x,y,t)I(x,y,t) + c(x,y,t)\Big) & \\ +e(x,y,t), & (x,y) \in \mathfrak{R}. \end{cases} \tag{5.24}$$

Similar to Equation (5.6), the gradient of the GDIM for both co-visible $\mathfrak{R}$ and occluded $\mathfrak{R}'$ regions can be written as

$$\begin{aligned} \nabla I(x,y,t) \;\; &= \Big(\nabla\Big(I(x-v_x dt, y-v_y dt, t-dt)\Big) + \nabla\Big(d(x,y,t)I(x,y,t)\Big) + \nabla c(x,y,t)\Big) \\ &\quad + r_1(x,y,t;dt) + r_2(x,y,t;dt). \end{aligned} \tag{5.25}$$

The term $\nabla I(x-v_x dt, y-v_y dt, t-dt)$ in (5.25) is nonlinear, it can be linearized as

$$\nabla I(x-v_x dt, y-v_y dt, t-dt) \approx \begin{cases} I_x(x,y,t) - I_{xx}(x,y,t)v_x(x,y,t) - I_{xy}(x,y,t)v_y(x,y,t), \\ I_y(x,y,t) - I_{yx}(x,y,t)v_x(x,y,t) - I_{yy}(x,y,t)v_y(x,y,t), \end{cases}$$

where I_{xx}, I_{xy}, I_{yx} and I_{yy} are second derivatives of the image brightness. Moreover,

$$\nabla\Big(d(x,y,t)I(x,y,t)\Big) = d(x,y,t)\nabla I(x,y,t) + I(x,y,t)\nabla d(x,y,t). \tag{5.26}$$

Using the linearized term and (5.26) in (5.25), we get:

$$\begin{pmatrix} I_{xx}v_x + I_{xy}v_y + I_{xt} - dI_x - Id_x - c_x \\ I_{yx}v_x + I_{yy}v_y + I_{yt} - dI_y - Id_y - c_y \end{pmatrix} = \begin{pmatrix} r_1 + r_2 \\ r_1 + r_2 \end{pmatrix}, \tag{5.27}$$

where d_x, c_x are horizontal, and d_y, c_y are vertical derivatives of brightness change parameters, such that $\nabla d = [d_x \; d_y]^{\mathrm{T}}$ and $\nabla c = [c_x \; c_y]^{\mathrm{T}}$. Note that we have omitted the spatiotemporal dependence of terms in (5.27). Following the same procedure as in Section 5.2.2, we arrive at the data term that incorporates the gradient of the GDIM:

$$E_{\text{data}}(\vec{v}, r_1) = \alpha||r_1||_1 + || \begin{pmatrix} I_{xx}v_x + I_{xy}v_y + I_{xt} - dI_x - Id_x - c_x \\ I_{yx}v_x + I_{yy}v_y + I_{yt} - dI_y - Id_y - c_y \end{pmatrix} - \begin{pmatrix} r_1 \\ r_1 \end{pmatrix} ||_2. \tag{5.28}$$

When the multiplier and the offset fields vary smoothly in the spatial domain, d_x, d_y, c_x and c_y are small; thus, these can be neglected. Neglecting Id_x, Id_y, c_x and c_y in (5.28) results in:

$$E_{\text{data}}(\vec{v}, r_1) = \alpha||r_1||_1 + || \begin{pmatrix} I_{xx}v_x + I_{xy}v_y + I_{xt} - dI_x \\ I_{yx}v_x + I_{yy}v_y + I_{yt} - dI_y \end{pmatrix} - \begin{pmatrix} r_1 \\ r_1 \end{pmatrix} ||_2. \tag{5.29}$$

A careful inspection of (5.29) reveals that the multiplier term d appears with the spatial image brightness derivatives I_x and I_y. A smoothly varying brightness change results in negligible values of I_x and I_y in low-textured video regions. Hence, dI_x and dI_y do not contribute much to the brightness constancy violation in smooth image regions. Image boundaries are likely to be associated with motion boundaries, and occlusion usually occurs at overlapping moving objects. Occluded regions also cause the brightness constancy violation at motion boundaries. The terms dI_x and dI_y may be neglected in low-textured image sequences. However, the contribution of these terms will be high for highly textured sequences and at image boundaries; therefore, these may not be neglected there. When image boundaries coincide with the occluded motion boundaries, it is safe to overlook dI_x and dI_y at image boundaries. Neglecting the terms dI_x and dI_y in (5.29), we get a data term that uses the brightness gradient constancy assumption (GCA):

$$E_{\text{data}}(\vec{v}, r_1) = \alpha||r_1||_1 + \quad || \begin{pmatrix} I_{xx}v_x + I_{xy}v_y + I_{xt} \\ I_{yx}v_x + I_{yy}v_y + I_{yt} \end{pmatrix} - \begin{pmatrix} r_1 \\ r_1 \end{pmatrix} ||_2. \tag{5.30}$$

Equation (5.30) can be written in the matrix form as

$$E_{\text{data}}(\mathbf{v}, \mathbf{r}_1) = \alpha||\mathbf{r}_1||_1 + || \begin{bmatrix} \text{diag}(\mathbf{I}_{xx}) & \text{diag}(\mathbf{I}_{xy}) & -\mathcal{I} \\ \text{diag}(\mathbf{I}_{yx}) & \text{diag}(\mathbf{I}_{yy}) & -\mathcal{I} \end{bmatrix} \begin{bmatrix} \mathbf{v}_x \\ \mathbf{v}_y \\ \mathbf{r}_1 \end{bmatrix} + \begin{bmatrix} \mathbf{I}_{xt} \\ \mathbf{I}_{yt} \end{bmatrix} ||_2, \tag{5.31}$$

where $\mathbf{I}_{xx}$, $\mathbf{I}_{xy}$, $\mathbf{I}_{yx}$ and $\mathbf{I}_{yy}$ are lexicographically vectorized second derivatives of the image brightness, and $\text{diag}(\mathbf{I}_{xx})$, $\text{diag}(\mathbf{I}_{xy})$, $\text{diag}(\mathbf{I}_{yx})$ and $\text{diag}(\mathbf{I}_{yy})$ are corresponding diagonal matrices, respectively. Let us define

$$C = \begin{bmatrix} \text{diag}(\mathbf{I}_{xx}) & \text{diag}(\mathbf{I}_{xy}) & -\mathcal{I} \\ \text{diag}(\mathbf{I}_{yx}) & \text{diag}(\mathbf{I}_{yy}) & -\mathcal{I} \end{bmatrix} \quad \text{and} \quad \mathbf{y}_2 = \begin{bmatrix} \mathbf{I}_{xt} \\ \mathbf{I}_{yt} \end{bmatrix}.$$

The data term integrating the occlusion term with the gradient constancy assumption is now given as

$$E_{\text{data}}(\mathbf{v}, \mathbf{r}_1) = \alpha||W\mathbf{r}_1||_1 + ||C\mathbf{u} + \mathbf{y}_2||_2, \tag{5.32}$$

where we have again replaced the ℓ_1 norm of $\mathbf{r}_1$ by the weighted ℓ_1 norm for a better approximation of the occlusion term as an indicator function. The following energy is then minimized for the joint optical flow estimation and occlusion detection:

$$\mathbf{u} = \arg\min_{\mathbf{v},\mathbf{r}_1} \alpha||W\mathbf{r}_1||_1 + ||C\mathbf{u} + \mathbf{y}_2||_2 + \lambda_1[\text{HVD}(\mathbf{v})]. \tag{5.33}$$

A comparison of (5.15) and (5.33) reveals that the regularization term remains the same. However, A and $\mathbf{y}_1$ have been replaced by C and $\mathbf{y}_2$ in the data term of (5.33). Optical flow estimation and occlusion detection by this modification of the proposed method can provide robustness against slowly varying brightness change patterns due to diffuse shading and specular flows.

5.4 Algorithmic details

In this section, we modify and then use a fast and first order algorithm NESTA, presented in [150], for the joint optical flow estimation and occlusion detection. The algorithmic details have been discussed for the minimization of the energy functional proposed in (5.23). The same algorithm can be applied for the modification of the proposed method

in (5.33) by letting λ_2 and λ_3 equal to zero in (5.23). We use differentiable Huber norm in place of the ℓ_1 norm for the occlusion term $W\mathbf{r}_1$ in (5.23). The regularization energy $E(\mathbf{x})$, to be minimized, can now be given as

$$E(\mathbf{x}) = \min_{\mathbf{v},\mathbf{r}_1,\mathbf{d},\mathbf{c}} \alpha||W\mathbf{r}_1||_\epsilon + ||B\mathbf{x}+\mathbf{y}_1||_2 + \lambda_1[\mathrm{HVD}(\mathbf{v})] + \lambda_2||\boldsymbol{\nabla}\mathbf{d}||_2 + \lambda_3||\boldsymbol{\nabla}\mathbf{c}||_2. \quad (5.34)$$

The algorithm solves the problem iteratively. The minimum of (5.34) at iteration k is computed as

$$\begin{aligned}\frac{\partial E(\mathbf{x}^k)}{\partial \mathbf{x}^k} = \partial_{\mathbf{x}^k} E(\mathbf{x}) = \alpha W^{\mathrm{T}} \frac{(W\mathbf{r}_1^k)}{\max(\epsilon,|W\mathbf{r}_1^k|)} + B^{\mathrm{T}}(B\mathbf{x}^k+\mathbf{y}_1) \\ +\lambda_1\Bigg(\boldsymbol{\nabla}_x^{\mathrm{T}} \frac{(\boldsymbol{\nabla}_x\mathbf{v}^k)}{\max(\epsilon,|\boldsymbol{\nabla}_x\mathbf{v}^k|)} + \boldsymbol{\nabla}_y^{\mathrm{T}} \frac{(\boldsymbol{\nabla}_y\mathbf{v}^k)}{\max(\epsilon,|\boldsymbol{\nabla}_y\mathbf{v}^k|)} + \boldsymbol{\nabla}_{xy}^{\mathrm{T}} \frac{(\boldsymbol{\nabla}_{xy}\mathbf{v}^k)}{\max(\epsilon,|\boldsymbol{\nabla}_{xy}\mathbf{v}^k|)} \\ +\boldsymbol{\nabla}_{yx}^{\mathrm{T}} \frac{(\boldsymbol{\nabla}_{yx}\mathbf{v}^k)}{\max(\epsilon,|\boldsymbol{\nabla}_{yx}\mathbf{v}^k|)}\Bigg) + \lambda_2\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{\nabla}\mathbf{d}^k + \lambda_3\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{\nabla}\mathbf{c}^k. \end{aligned} \quad (5.35)$$

The algorithm computes two auxiliary variables $\mathbf{p}^k$ and $\mathbf{q}^k$ at each iteration from $\partial_{\mathbf{x}^k} E(\mathbf{x}^k)$, and combines them to get the next estimate of the parameter vector $\mathbf{x}^k$. Lipschitz constant L is required for the computation of auxiliary variables in the proposed algorithm. Lipschitz constant depends on α, λ_1, λ_2, λ_3, Huber norm parameter ϵ and norms of B and TV, which can be given as

$$L = \max(\frac{16\lambda_1,\alpha}{\epsilon},\lambda_2,\lambda_3) + ||B^{\mathrm{T}}B||_2. \quad (5.36)$$

The proposed algorithm is applied in a coarse-to-fine image sequence pyramid to handle large displacements of pixels. A Laplacian image sequence pyramid is constructed which contains consecutive video frames at different resolutions for a coarse-to-fine estimation of optical flow. Sparse occlusion term $\mathbf{r}_1$, optical flow $\mathbf{v}$, and brightness change parameters $\mathbf{d}$ and $\mathbf{c}$ are initialized at 0 in the coarsest level of the pyramid. At each level of the pyramid, the detected occlusion, the estimated optical flow and estimates of these parameters are propagated to the next finer level as initial guesses. The algorithm runs for a fixed number of iterations at each pyramid level or until it reaches the convergence. The summary of the algorithm is given in Table 5.1.

Table 5.1: The summary of the algorithm for the joint optical flow estimation and occlusion detection.

Initialization: $\mathbf{x}^0 = [\mathbf{v}_x^0 \;\; \mathbf{v}_y^0 \;\; \mathbf{r}_1^0 \;\; \mathbf{d}^0 \;\; \mathbf{c}^0]^{\mathrm{T}} = 0$,
for level=pyramid_levels, ... , 1
  if(level= pyramid_levels)
    $\mathbf{x} = \mathbf{x}^0$,
  else
    $\mathbf{x}^0 =$ up-sample($\mathbf{x}$),
  end
  Set iteration index $k = 1$,
  while(not converged & $k \leq$ max_iter)
    1. compute $\partial_{\mathbf{x}^k} E(\mathbf{x}^k)$ from Equation (5.35),
    2. compute $\gamma^k = \frac{1}{2}(k+1)$ and $\tau^k = \frac{2}{k+3}$,
    3. compute $\mathbf{p}^k = \mathbf{x}^k - \frac{1}{L}\partial_{\mathbf{x}^k} E(\mathbf{x}^k)$,
    4. compute $\mathbf{q}^k = \mathbf{x}^0 - \frac{1}{L}\sum_i^k \gamma^i \partial_{\mathbf{x}^i} E(\mathbf{x}^i)$,
    5. update $\mathbf{x}^k = \tau^k \mathbf{p}^k + (1-\tau^k)\mathbf{q}^k$,
    $k = k+1$,
  end while
  $\mathbf{x} = \mathbf{x} + \mathbf{x}^0$.
end for

5.5 Experimental Results and Analysis

In this section, the performance of the proposed method is evaluated and experimental results are analyzed. The experimental setup and datasets used in experiments are described in Section 5.5.1. The performance of the proposed method is assessed and compared with other methods on three different kinds of brightness change scenarios: diffuse shading, specular, and the combined diffuse shading and specular flows. Therefore, three different sets of experiments are conducted on video sequences undergoing brightness change due to diffuse shading only in Section 5.5.2, specular only in Section 5.5.3 and the combined diffuse shading and specular flows in Section 5.5.4. Finally, the proposed method have been applied to challenging real sequences where multiple factors cause the brightness change in Section 5.5.5.

5.5.1 Experimental Setup and Video Datasets

The proposed method has been implemented in a coarse-to-fine pyramid with a down-sampling factor of 0.70. A maximum of 10 pyramid levels are used to handle large displacements. It is found that values of $(\alpha, \lambda_1) \in \{0.01, ..., 1\}$ and $(\lambda_2, \lambda_3) \in \{0.1, ..., 10\}$ give reasonable accuracy; thus, these parameters are manually tuned to get the best results in terms of small optical flow estimation error and high occlusion detection precision. The value of ϵ in the ℓ_1 norm approximation plays an important role in the convergence of the proposed algorithm. A value of $\epsilon = 0.01$ is used in all of the experiments which gives reasonable accuracy with good convergence rate. The proposed algorithm converges in a few hundred iterations; therefore, it is run for a maximum of 1500 iterations or until it reaches the convergence.

In all of the experiments, the performances of the proposed method and its modified version, referred to as **proposed w/o PARA**, have been compared with the brightness constancy based sparse occlusion detection method of Ayvaci et al. [5]. Since pre-filtering of video sequences provides robustness against the brightness change, the aforementioned sparse occlusion detection method is run with video sequences preprocessed by a high pass filter. This method has been referred to as **PreFilt Ayvaci et al.** in the experiments; it has also been compared with the proposed method.

These experiments are conducted on video sequences taken from publicly available datasets. Middlebury video sequence dataset is one of the most famous to evaluate performances of optical flow estimation methods. However, this dataset does not provide the ground-truth for occluded regions. To compare performances of occlusion detection methods, video sequences of Middlebury have been manually labeled either visible or occluded in [5]. Labeled videos of this dataset are used here to assess and compare the performance of the proposed method with the aforementioned methods. The values of $\alpha = 0.01$, $\lambda_1 = 0.2$, $\lambda_2 = 5$ and $\lambda_3 = 5$ have been used for video sequences of Middlebury dataset. The aforementioned methods are also tested on real video sequences which have been taken from the dataset produced by Stein and Hebert [6]. This dataset contains

indoor and outdoor scenes; it has been used for occlusion and object boundary detection [170]. The presence of image noise, artefacts due to compression, and abrupt camera motions makes this dataset quite challenging. The values of $\alpha = 0.004$, $\lambda_1 = 0.1$, $\lambda_2 = 5$ and $\lambda_3 = 5$ have been used for the real sequences of this dataset.

Middlebury provides the ground-truth optical flow for training videos; however, these video sequences do not violate the brightness constancy assumption. To analyze performances of aforementioned methods under brightness changes due to diffuse shading, specular, and the combined diffuse shading and specular flows, a controlled magnitude of violation is used in these video sequences. A smoothly varying Gaussian function G is used to model smooth brightness change. Diffuse shading and specular flows are modeled, respectively, as the multiplier and additive fields by using this function. The magnitude of this function is changed to get different amounts of brightness changes. Since the function G has values in the [0 1] range, the brightness (intensity) of video sequences is also normalized to the values in the same range. The brightness change $\Delta\mathbf{I}$ in any video frame, due to the function G, gives a measure of the brightness constancy violation. Here, the normalized brightness change is defined as the mean of the ratio of the brightness change to the original brightness as

$$\text{Normalized brightness change (NBC)} := \text{mean}(|\delta\mathbf{I}|/\mathbf{I}),$$

where the mean is taken over the entire image domain. A high value of the NBC represents severe violation of the brightness constancy and vice versa. A value of NBC = 0 means that the brightness constancy assumption is satisfied. The magnitude of the brightness change is gradually increased in the subsequent video frame, to be matched with the reference frame, by increasing the value of the NBC. The aforementioned methods are then applied on the resulting sequences undergoing brightness changes for the comparison of their performances.

Separate evaluation measures for the estimated optical flow and the detected occlusion are provided. Optical flow estimation results are reported in terms of the mean end point

error (MEPE) for video sequences with known ground-truth. The MEPE is given as

$$\text{MEPE} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{[(\mathbf{v}_x)_i - (\mathbf{v}_{x\text{gt}})_i]^2 + [(\mathbf{v}_y)_i - (\mathbf{v}_{y\text{gt}})_i]^2}, \tag{5.37}$$

where $\mathbf{v}_{\text{gt}} = [\mathbf{v}_{x\text{gt}}\ \mathbf{v}_{y\text{gt}}]^{\text{T}}$ is the ground-truth optical flow. The MEPE is drawn against the NBC for a quantitative comparison of estimated optical flows. Furthermore, precision recall (PR) curves are drawn against the NBC to measure the performances of the aforementioned methods to detect occlusion. Different color schemes are also used to show the estimated optical flow and the detected occlusion for visual comparison of their quality.

5.5.2 Violation due to Diffuse Shading Flow

The first set of experiments are carried out on aforementioned methods when the brightness constancy violation is only due to diffuse shading flow. Figure 5.3 shows the mean end point error results against the NBC for methods: Ayvaci et al. [5], PreFilt Ayvaci et al., the proposed method and the proposed w/o PARA. In this figure $\log_{10}$ (MEPE) is plotted for a better display. These results have been computed on eight training sequences of Middlebury whose ground-truth is known. An average of the MEPE is taken over all of these sequences for each of the method. A sharp rise in the MEPE can be observed for increasing values of the NBC by the method of Ayvaci et al. The results by PreFilt Ayvaci et al. do not show much increase in the error; however, this method results in larger values of MEPE than the proposed method as the NBC increases. The MEPE remains almost constant in the proposed method and the proposed w/o PARA even for large values of NBC.

Figure 5.4 presents the estimated optical flows shown in color by these methods. These results have been obtained when the target video frame, to be matched with the reference video frame, exhibits a brightness constancy violation corresponding to the NBC = 0.5. Figure 5.4 (b) shows that the method of Ayvaci et al. fails to estimate optical flow correctly, thereby producing erroneous flow vectors especially in the highlighted part of the video sequence. The performance of PreFilt Ayvaci et al. is comparatively better than Ayvaci et al.; however, a slight blurring of motion boundaries can be observed which

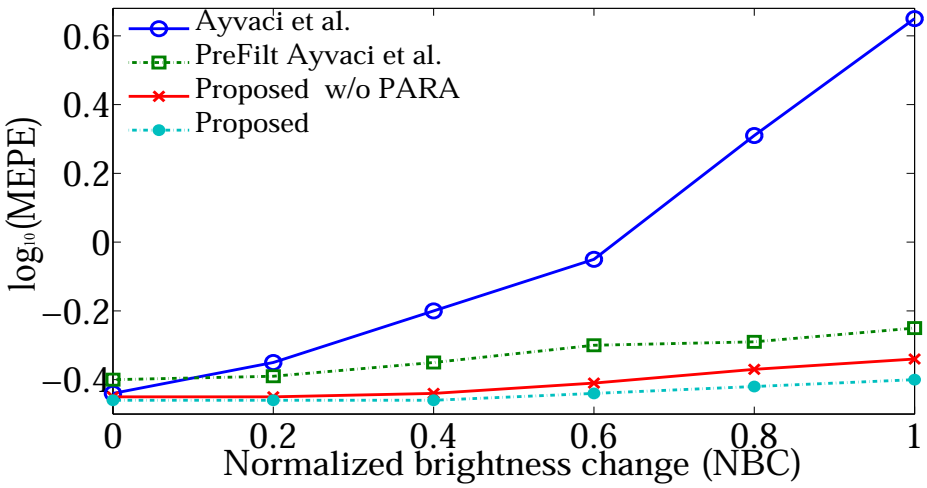


Figure 5.3: Optical flow estimation results computed on eight training sequences taken from Middlebury under brightness constancy violation caused by diffuse shading flow only. The $\log_{10}$ (MEPE) is given as a function of the NBC for all methods.

(a) Ground-truth (b) Ayvaci et al. (c) PreFilt Ayvaci et al. (d) Proposed (e) Proposed w/o PARA

Figure 5.4: The estimated optical flow, displayed in color, in Middlebury *Grove3* sequence. (a) The ground-truth optical flow and the results by methods: (b) Ayvaci et al. [5], (c) PreFilt Ayvaci et al., (d) the proposed method and (e) the proposed w/o PARA. These results have been reported for a diffuse shading flow violation corresponding to NBC = 0.5. The color scheme used to display the estimated optical flow is also shown.

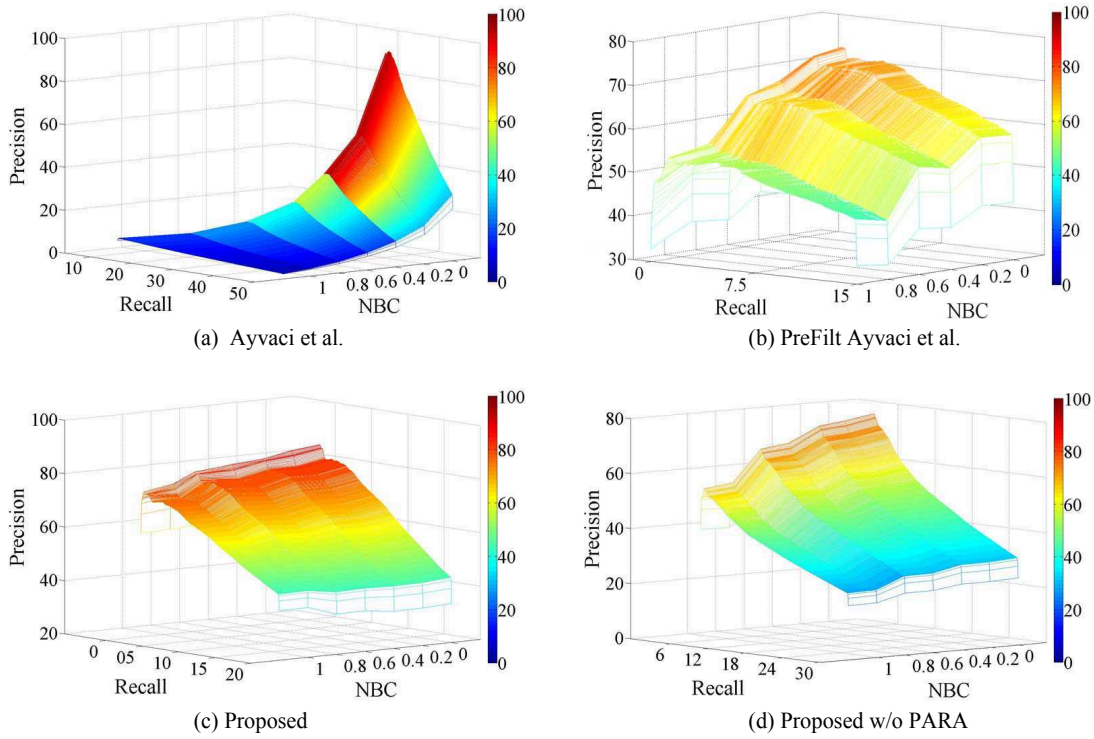
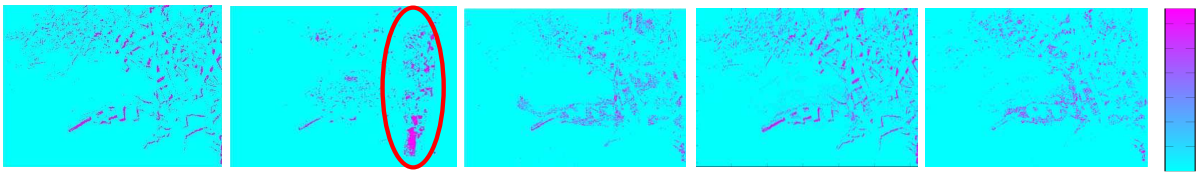


Figure 5.5: Occlusion detection results computed on eight training sequences of Middlebury under brightness constancy violation caused by diffuse shading flow only. PR curves as functions of the NBC for methods: (a) Ayvaci et al. [5], (b) PreFilt Ayvaci et al., (c) the proposed method and (d) the proposed w/o PARA.

results in an increased MEPE in this method. Figure 5.4 (d) and (e), respectively, show that estimated optical flows by proposed method and proposed w/o PARA are close to ground-truth flow,

Figure 5.5 shows precision recall (PR) curves for the comparison of occlusion detection performances of aforementioned methods. These curves have been drawn against a set of increasing NBC. Each PR curve shows an average of the results computed on eight training sequences of Middlebury. These curves are smoothed by the use of a Gaussian filter for a better display. Figure 5.5 (a) shows that the precision of the method of Ayvaci et al. decreases rapidly as the NBC increases. PreFilt Ayvaci et al. and the proposed method, on the other hand, perform much better than Ayvaci et al. for relatively higher values of the NBC. However, there is a decrease in the occlusion detection precision of



(a) Ground-truth (b) Ayvaci et al. (c) PreFilt Ayvaci et al. (d) Proposed (e) Proposed w/o PARA

Figure 5.6: The detected occlusion in Middlebury *Grove3* sequence. (a) The ground-truth occlusion and the results by methods: (b) Ayvaci et al. [5], (c) PreFilt Ayvaci et al., (d) the proposed method and (e) the proposed w/o PARA for the NBC = 0.5 under diffuse shading flow violation.

PreFilt Ayvaci et al. for higher values of the NBC. Results by the proposed method and the proposed w/o PARA in Figure 5.5 (c) and (d), respectively, indicate that both perform superior to PreFilt Ayvaci et al. for the task of occlusion detection under severe violation of the brightness constancy.

Figure 5.6 presents the detected occlusion shown in color for a value of NBC = 0.5. The results in Figure 5.6 (b) show that the method of Ayvaci et al. wrongly detects occlusion. Moreover, PreFilt Ayvaci et al. fails to detect some of occluded regions. This happens when the strong dominant violation across the motion boundaries results in the wrong matching of scene points of the reference video frame with occluded points of the consecutive video frame. Figure 5.6 (e) shows that the proposed w/o PARA also shows some true negatives. Moreover, the performance of the proposed method in Figure 5.6 (d) is better than the rest of the methods.

5.5.3 Violation due to Specular Flow

The second set of experiments are conducted on video sequences when specular flow causes the brightness constancy violation. Figure 5.7 shows the mean end point error results against increasing values of the NBC. In this figure $\log_{10}$(MEPE) is plotted for a better display. The MEPE increases for increasing values of the NBC by the method of Ayvaci et al. Unlike the first set of experiments, there is a less rise in the MEPE by this method under specular flow violation. This is due to the image smoothing by a small Gaussian filter in coarse-to-fine image pyramid construction. This filter is applied

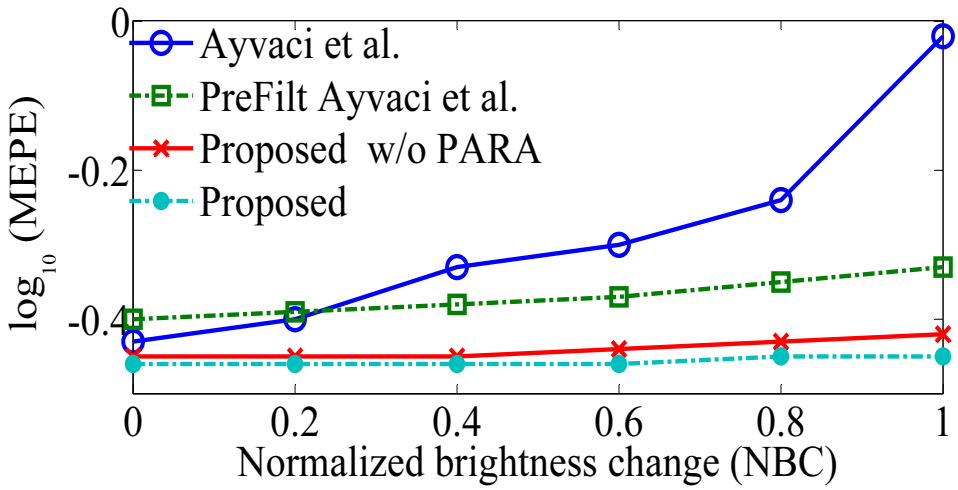


Figure 5.7: Optical flow estimation results computed on eight training sequences of Middlebury under brightness constancy violation caused by specular flow only. The $\log_{10}$ (MEPE) is given as a function of the NBC for all methods.

to raw video sequences to avoid aliasing in the resultant low resolution video frames. This filtering provides some robustness against specular flow violation. The proposed method does not show much increase in the error for higher values of the NBC, and its performance is better than PreFilt Ayvaci et al. specially for higher values of the NBC.

To assess the occlusion detection performance of all four methods, precision recall curves have been drawn against increasing values of the NBC (see Figure 5.8). A rapid decrease in the precision of the method of Ayvaci et al. can be seen in Figure 5.8 (a) as the NBC increases. PreFilt Ayvaci et al. and the proposed method demonstrate the capability to detect occlusion in video regions undergoing the brightness constancy violation. However, the precision of the detected occlusion decreases in the former method as the value of the NBC increases. For small values of the NBC, observe that precisions of the proposed method and the proposed w/o PARA in Figure 5.8 (c) and (d) are more than the precision of PreFilt Ayvaci et al., as shown in Figure 5.8 (b).

Figure 5.9 and Figure 5.10, respectively, show estimated optical flows and detected occlusions for aforementioned methods. These figures report the results for a value of NBC = 0.5. The method of Ayvaci et al. estimates flow erroneously as shown in the

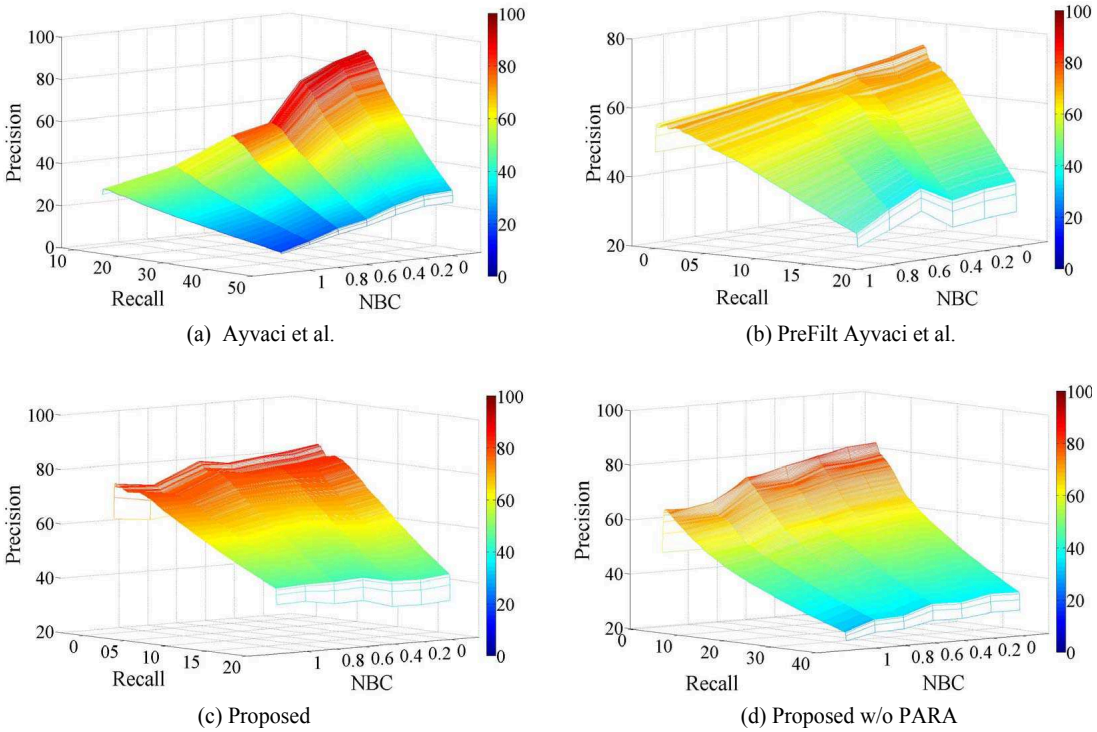


Figure 5.8: Occlusion detection results computed on eight training sequences of Middlebury under brightness constancy violation caused by specular flow only. PR curves as functions of the NBC for methods: (a) Ayvaci et al. [5], (b) PreFilt Ayvaci et al., (c) the proposed method and (d) the proposed w/o PARA.

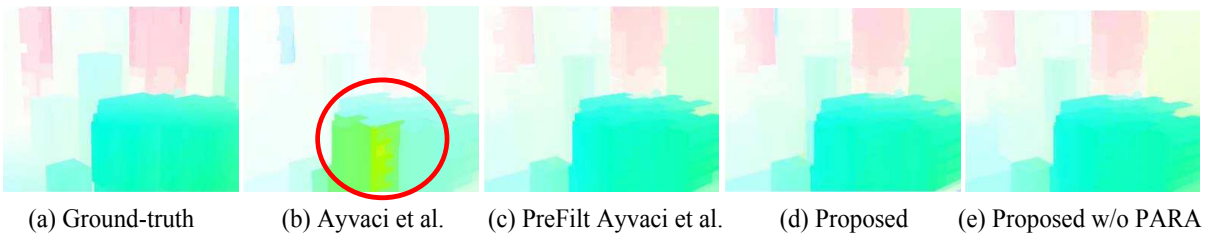


Figure 5.9: The estimated optical flow, displayed in color, in Middlebury *Urban2* sequence. (a) The ground-truth optical flow and the results by methods: (b) Ayvaci et al. [5], (c) PreFilt Ayvaci et al., (d) the proposed method and (e) the proposed w/o PARA. These results have been reported for a specular flow violation corresponding to NBC = 0.5.

(a) Ground-truth (b) Ayvaci et al. (c) PreFilt Ayvaci et al. (d) Proposed (e) Proposed w/o PARA

Figure 5.10: The detected occlusion in Middlebury *Urban2* sequence. (a) The ground-truth occlusion and the results by methods: (b) Ayvaci et al. [5], (c) PreFilt Ayvaci et al., (d) the proposed method and (e) the proposed w/o PARA, for NBC = 0.5 under specular flow violation.

highlighted part of Figure 5.9 (b). The erroneous flow results in the wrong detection of occluded video regions as shown in Figure 5.10 (b). The proposed method detects occlusion better than PreFilt Ayvaci et al., see Figure 5.10 (c) and (d) for the comparison. Observe that some false positives are detected by the proposed method w/o PARA as shown in the highlighted part of Figure 5.10 (e). However, these results are better than the results obtained by PreFilt Ayvaci et al.

5.5.4 Combined Diffuse Shading and Specular Violation

In the third set of experiments, the performances of the methods mentioned in Section 5.5.1 have been evaluated when both diffuse shading and specular flows contribute to the brightness constancy violation. In the combined violation, although different values of the NBC can be used for multiplicative and additive functions, we have used same values of the NBC for both brightness change functions. In these experiments, a value of NBC = 0.5 represents that both diffuse shading and specular flows are contributing equally to the violation of brightness constancy.

Figure 5.11 presents the mean end point error as a function of the NBC. In this figure $\log_{10}$(MEPE) is plotted for a better display. A rapid rise in the MEPE can be seen for increasing values of the NBC by the method of Ayvaci et al. Results by PreFilt Ayvaci et al. also show an increase in the MEPE which is much lesser than the increase in the MEPE observed by Ayvaci et al. The MEPE by the proposed method, on the other hand, remains almost constant. Moreover, optical flow estimation error performance by

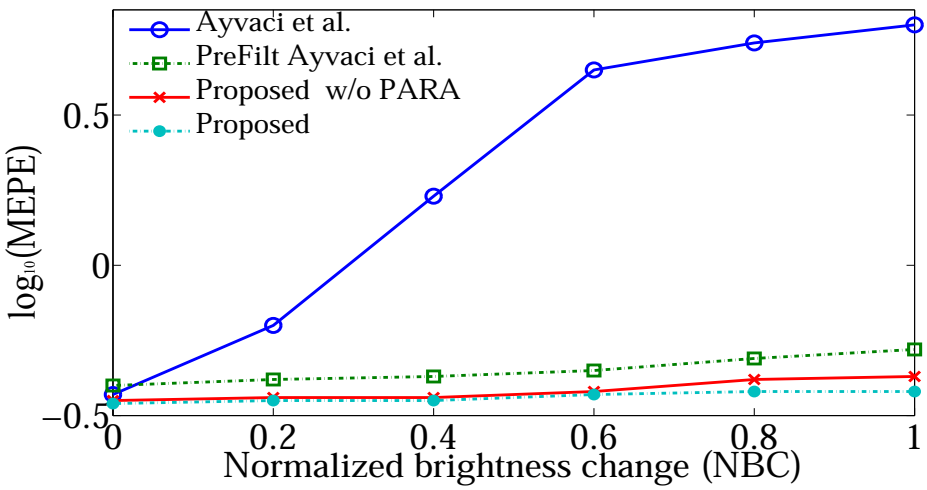


Figure 5.11: Optical flow estimation results computed on eight training sequences of Middlebury under brightness constancy violation caused by combined diffuse shading and specular flows. The $\log_{10}$ (MEPE) is given as a function of the NBC for all methods.

the proposed method is slightly better than the proposed w/o PARA for higher values of the NBC.

The estimated optical flows by the evaluated methods have been presented in Figure 5.12. The first and second rows show results on *Urban2* and *Grove3* sequences, respectively. These result have been reported for a value of NBC = 0.5. The method of Ayvaci et al. produces wrong optical flow estimates especially in highlighted parts of video sequences as shown in Figure 5.12 (b). These parts of videos undergo severe brightness constancy violation. Highlighted parts of both video sequences in Figure 5.12 (c) show erroneous flow vectors by PreFilt Ayvaci et al. The quality of estimated optical flows, produced by the proposed method and the proposed w/o PARA in Figure 5.12 (d) and (e), respectively, is better than other methods.

Figure 5.13 shows PR curves against different values of the NBC by evaluated methods. A rapid decrease in the precision can be noticed by Ayvaci et al. in Figure 5.13 (a). The precision of PreFilt Ayvaci et al. also decreases when the value of the NBC increases as shown in Figure 5.13 (b). Again, the proposed method demonstrates the capability to detect occlusion in video regions significantly violating the brightness constancy as shown

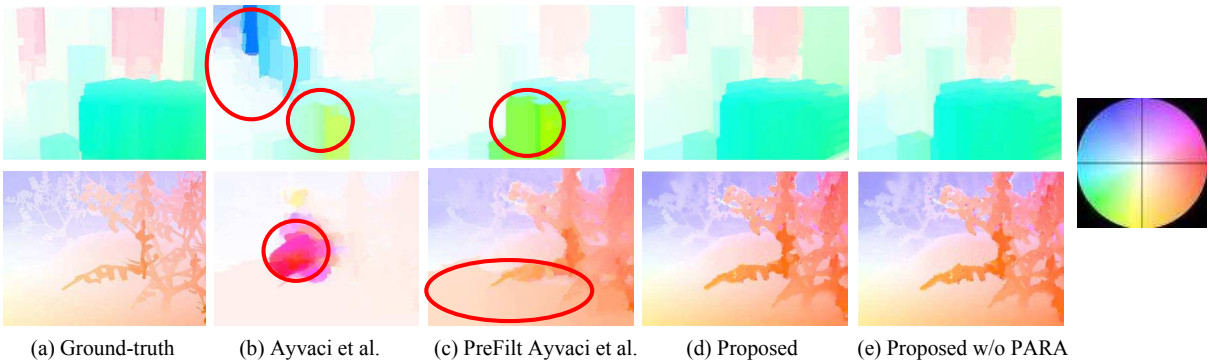


Figure 5.12: Estimated optical flows in *Urban2* (top) and *Grove3* (bottom) sequences, displayed in color. (a) The ground-truth flow and results by methods: (b) Ayvaci et al. [5], (c) PreFilt Ayvaci et al., (d) the proposed method and (e) the proposed w/o PARA. These results have been reported for a combined diffuse shading and specular flows violation corresponding to NBC = 0.5.

in Figure 5.13 (c). However, a small degradation in the occlusion detection performance by the proposed w/o PARA can also be observed in Figure 5.13 (d) for large values of the NBC.

Figure 5.14 shows detected occlusions by all four methods. Video regions violating the brightness constancy have been wrongly detected as occluded regions by the method of Ayvaci et al., as shown in Figure 5.14 (b). Moreover, the erroneous optical flow estimates result in the wrong occlusion detection in PreFilt Ayvaci et al. as shown in the highlighted part of Figure 5.14 (c). The quality of the detected occlusion is better by the proposed method than other methods as shown in Figure 5.14 (d).

5.5.5 Results on Real Sequences

In these experiments, aforementioned methods have been applied to real video sequences for the comparison of their performances. Real sequences have been taken from the occlusion boundary detection dataset developed by Stein and Hebert [6]. Here, the results have been reported on an indoor sequence *Mugs2* and an outdoor sequence *Fencepost*. *Mugs2* sequence has been captured by a moving camera. This sequence does not undergo brightness change. In *Fencepost* sequence, shadows of tree leaves on the fencepost cause

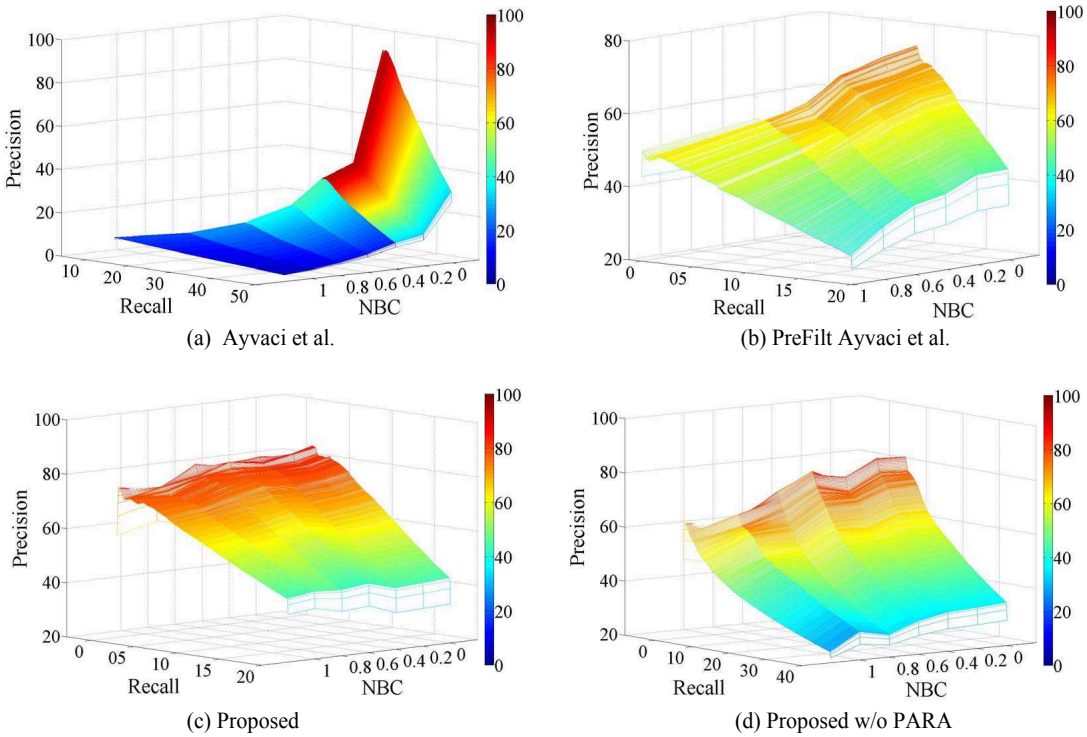


Figure 5.13: Occlusion detection results computed on eight training sequences of Middlebury under brightness constancy violation caused by combined diffuse shading and specular flows. PR curves as functions of the NBC for methods: (a) Ayvaci et al. [5], (b) PreFilt Ayvaci et al., (c) the proposed method and (d) the proposed w/o PARA.

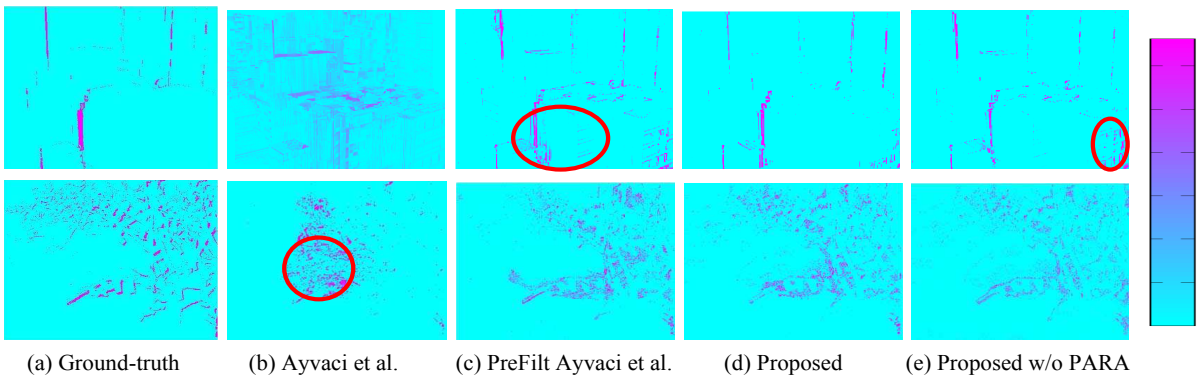


Figure 5.14: Occlusion detection in *Urban2* (top) and *Grove3* (bottom) sequences. (a) The ground-truth occlusion and results by methods: (b) Ayvaci et al. [5], (c) PreFilt Ayvaci et al., (d) the proposed method and (e) the proposed w/o PARA. These results have been reported for a combined diffuse shading and specular flows violation corresponding to the NBC = 0.5.

the violation of the brightness constancy assumption. It is worth mentioning that, in these experiments, the controlled magnitude of violation is not added to the sequences. Since this dataset provides the ground-truth for occlusion boundaries only, the quality of the estimated optical flow is assessed by reconstructing the target video frame using estimated optical flow. The mean square error (MSE) between the original target frame and the reconstructed target frame is then reported. The reference video frame is warped using the estimated optical flow to reconstruct the target video frame.

Figure 5.15 shows estimated optical flows in aforementioned video sequences. The method of Ayvaci et al. produces wrong optical flow estimates in the highlighted part of *Fencepost* sequence as shown in Figure 5.15 (a). The MSE between the original target and the reconstructed target video frames has been presented in Figure 5.16. Results on *Fencepost* sequence in Figure 5.16 (a) show that the proposed method performs better than Ayvaci et al. and PreFilt Ayvaci et al. in terms of the MSE. *Mugs2* sequence does not show the violation of the brightness constancy assumption; therefore, the performance of the method of Ayvaci et al. is comparable to all other methods in terms of the MSE as shown in Figure 5.16 (b).

Figure 5.17 shows detected occlusions for all four methods computed on *Fencepost* (top row) and *Mugs2* (bottom row) sequences. The highlighted part of Figure 5.17 (b) shows a lot of false positives produced by the method of Ayvaci et al. in *Fencepost* sequence. PreFilt Ayvaci et al. performs slightly better than the method without pre-filtering. The proposed w/o PARA results in some true negatives in the right half of *Fencepost* sequence as shown in Figure 5.17 (e). However, the proposed method performs comparatively better than all other methods as shown in Figure 5.17 (d). Results on *Mugs2* sequence have been presented in the bottom row of Figure 5.17. Ayvaci et al. and PreFilt Ayvaci et al. in Figure 5.17 (b) and (c), respectively, fail to detect a lot of occluded pixels in *Mugs2* sequence. The proposed w/o PARA introduces some false positives specially in the upper right portion of the video as shown in the highlighted part of Figure 5.17 (e). This is because the gradient of the GDIM highlights the edges in the scene, which could result

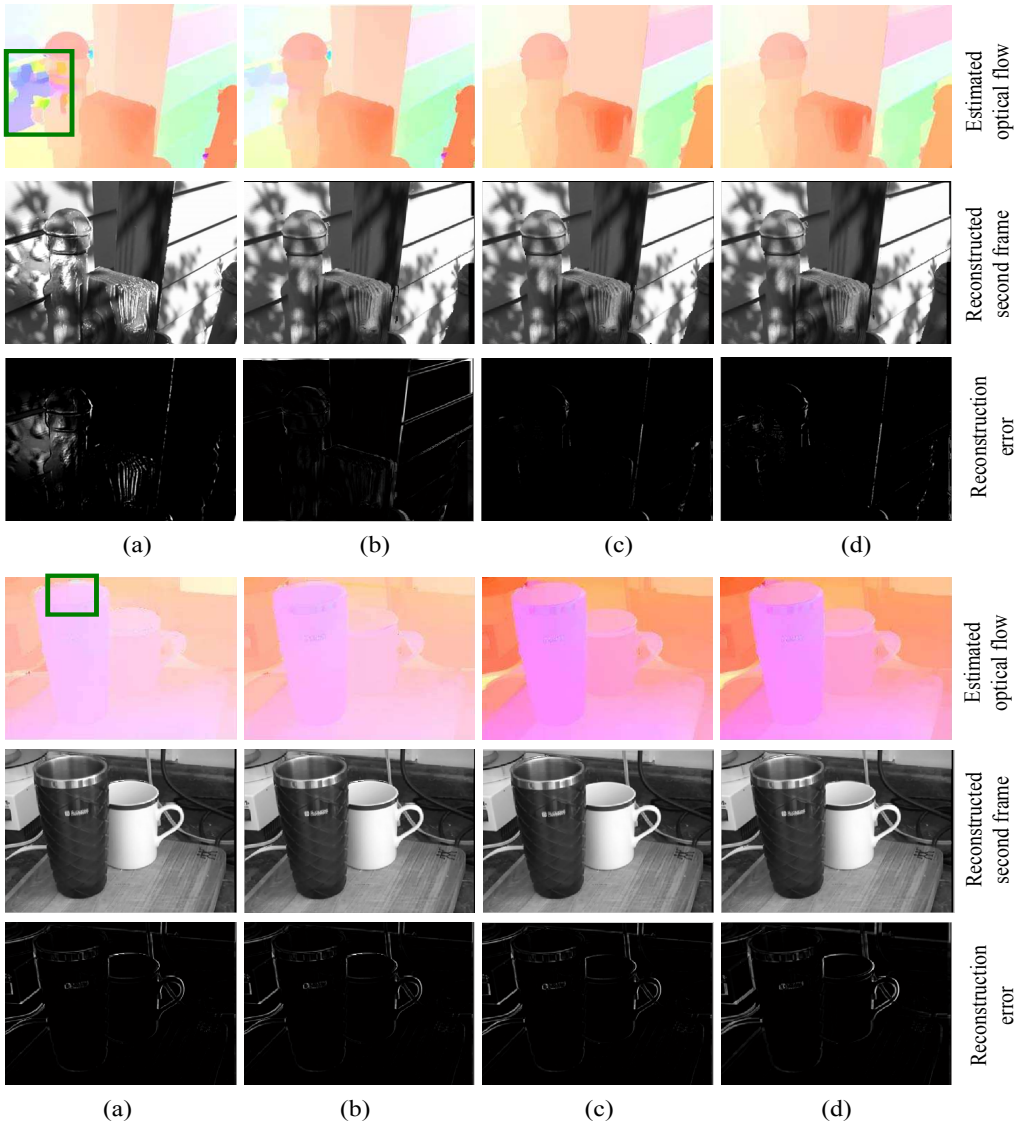


Figure 5.15: Optical flow estimation on video sequences *Fencepost* (top) and *Mugs2* (bottom) by (a) Ayvaci et al. [5], (b) PreFilt Ayvaci et al., (c) the proposed method and (d) the proposed w/o PARA. The estimated optical flow is shown in the 1st row while the 2nd row shows the target video frame which is reconstructed by warping the reference video frame using the estimated optical flow, and the visualization of the reconstruction error (3rd row).

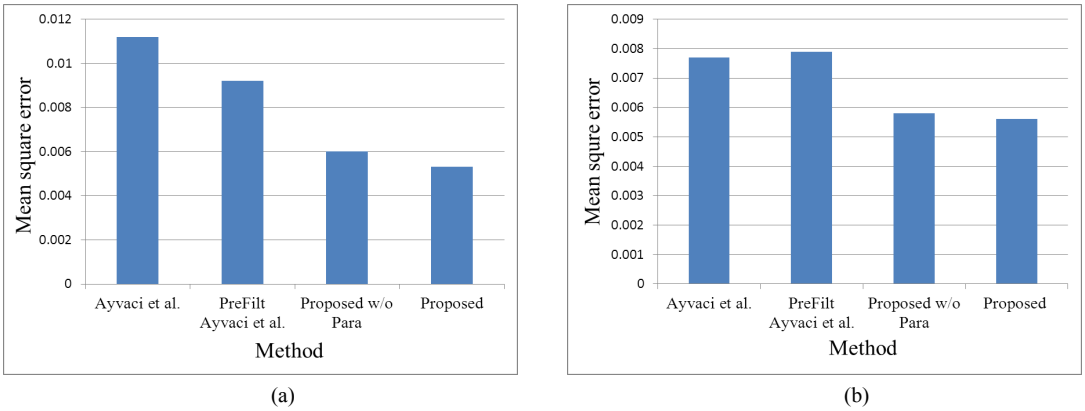


Figure 5.16: The mean square reconstruction error between the original target video frame and the reconstructed target frame computed on real sequences (a) *Fencepost* and (b) *Mugs2* taken from [6].

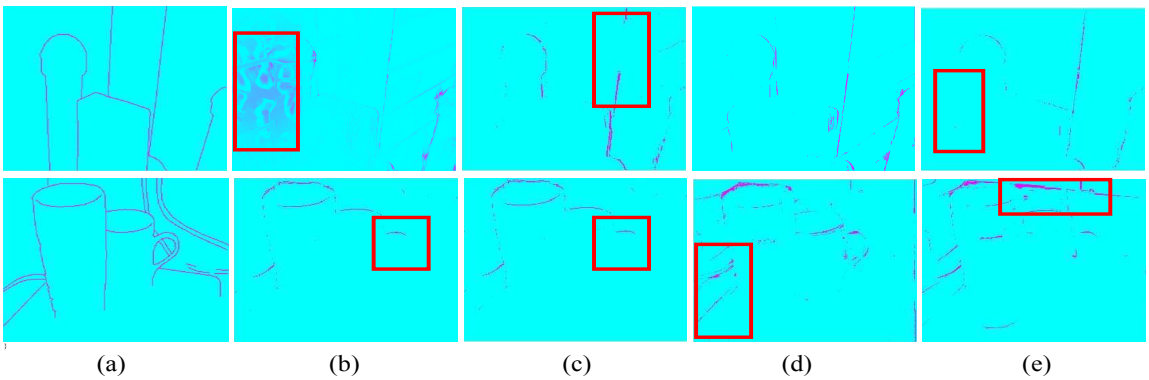


Figure 5.17: Occlusion detection results on *Fencepost* (top) and *Mugs2* (bottom) sequences, (a) ground-truth occlusion, occlusions detected by (b) Ayvaci et al. [5], (c) PreFilt Ayvaci et al. [5], (d) the proposed method and (e) the proposed w/o PARA.

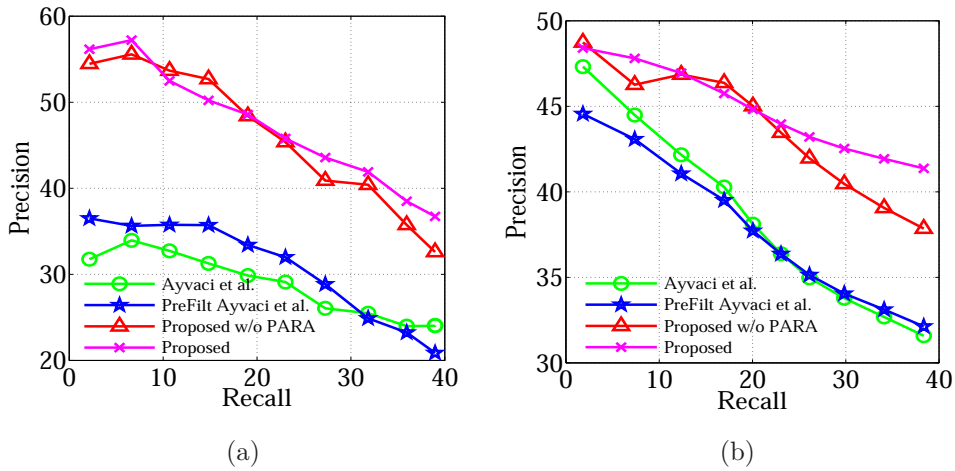
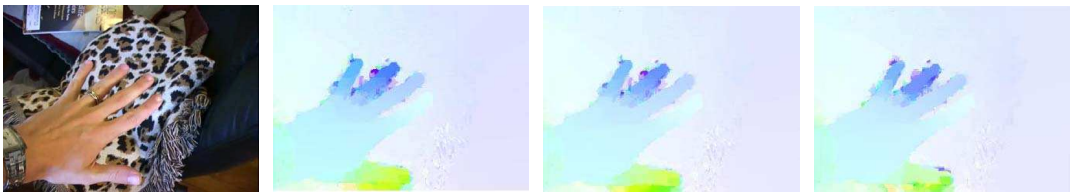


Figure 5.18: The precision recall curves for the detected occlusions by methods: Ayvaci et al. [5], PreFilt Ayvaci et al., the proposed method and the proposed w/o PARA. These results are reported on (a) *Fencepost* and (b) *Mugs2* sequences.

in false positives when the image boundaries do not coincide with motion boundaries. Figure 5.17 (d) shows that the performance of the proposed method is better than other methods.

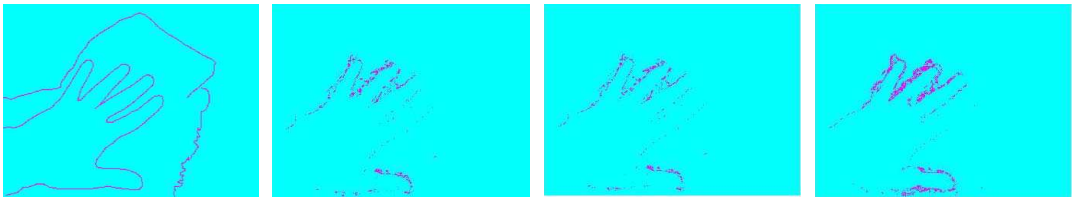
Figure 5.18 shows precision recall (PR) curves on both of the sequences. PR curves for *Fencepost* sequence in Figure 5.18 (a) show that the occlusion detection performance of the method of Ayvaci et al. and PreFilt Ayvaci et al. is lower than both the proposed method and its modified version. Moreover, when compared to the proposed method, there is a decrease in the performance of these two methods for *Mugs2* sequence as shown in Figure 5.18 (b).

It is found that the occlusion detection performance of the proposed optical flow based method decreases for slowly moving objects in these sequences. Therefore, the proposed method performs poor on video sequences which contain slowly moving objects. One of the examples is *Hand3* sequence, where a human hand moves abruptly in front of a slowly moving cloth. Figure 5.19 shows estimated optical flows by aforementioned methods on this sequence. A lot of outliers in the estimated optical flow can be observed around fast movement of the fingers. This problem is inherent to coarse-to-fine optical flow estimation



(a) (b) (c) (d)

Figure 5.19: The estimated optical flow on real sequence *Hand3* taken from [6]. (a) The reference video frame, estimated optical flows by (b) the method of Ayvaci et al., (c) PreFilt Ayvaci et al. and (d) the proposed method. Note that all methods produce erroneous flow vectors across swiftly moving fingers.



(a) (b) (c) (d)

Figure 5.20: The detected occlusion on real sequence *Hand3* taken from [6]. (a) The ground-truth occlusion, the detected occlusions by the methods: (b) Ayvaci et al., (c) PreFilt Ayvaci et al. and (d) the proposed method. Note that all methods are unable to detect occluded pixels associated with the slowly moving cloth behind abruptly moving human hand.

methods; thus, these methods produce erroneous flow vectors at the boundaries of fast moving hand. This, in turn, results in wrong detection of occluded pixels. Observe that all methods are unable to detect occluded pixels associated with the slowly moving cloth as shown in Figure 5.20 (b)-(d). This degradation in the occlusion detection performance is attributed to the image smoothing in the construction of image pyramid. Since the proposed sparse occlusion detection method uses the residual of its data term, a very small movement looks even smaller on a coarser level of the pyramid. Therefore, the smoothing of video sequences in coarse-to-fine image pyramid construction blurs sharp image boundaries. The data term wrongly matches occluded pixels to visible pixels when motion boundaries correspond to image boundaries. Due to this wrong matching of occluded pixels, the method of Ayvaci et al. and the proposed method detect these pixels as visible.

Finally, the run time of the proposed method is compared with its modified version. It can be seen in Figure 5.21 that there is not much difference between the time taken by the proposed method and the proposed w/o PARA for small-sized video frames. However, this difference grows as the size of the video frame increases.

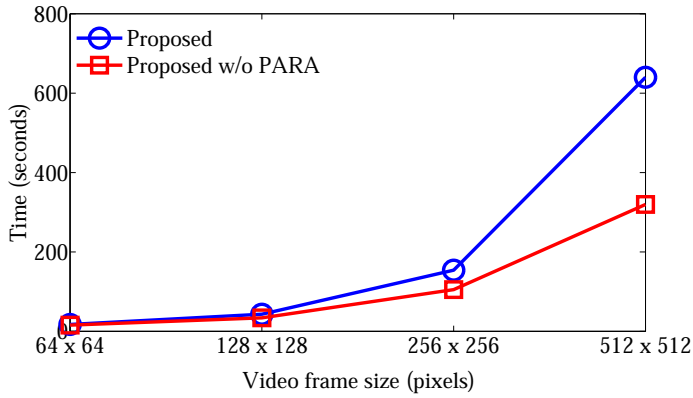


Figure 5.21: The run time of the proposed method and its modified version, proposed w/o PARA, as a function of the size of the video frame.

5.6 Discussion and Conclusions

This chapter proposes methods to detect sparse occlusion in digital videos using optical flow under brightness constancy violation. Occluded video regions and varying illumination of the scene are modeled separately. Occluded video regions are generally sparse in any digital video; therefore, a sparsity constraint is imposed on the occlusion term in a variational energy minimization framework. Moreover, a generalized dynamic image model (GDIM) is used to model the brightness constancy violation. This model captures the brightness constancy violation caused by diffuse shading and specular flows. A modification of the proposed method is also presented, which does not require the estimation of brightness change parameters of the GDIM under certain circumstances. These circumstances have also been discussed, under which the modification of the proposed method is applicable.

The presented method uses cues from the residual of the intensity matching term to detect sparse occlusion. This residual is usually small when the motion of objects is small.

It is found that the proposed method can successfully detect occlusion when there is a sufficient interframe motion of objects. However, the precision of this method, to detect occluded video regions, decreases in case of very small interframe motion. This degradation in the occlusion detection performance is attributed to coarse-to-fine image pyramid construction in optical flow estimation. Static occlusion cues, for instance, brightness, color and texture have been found to work better in this case. These cues can be integrated with dynamic cues to enhance occlusion detection capabilities of optical flow estimation methods as proposed in [170].

Chapter 6

Summary and Future Directions

Chapter contents

This thesis has been dedicated to the design of sparse variational regularization for visual motion estimation in digital videos. The work presented in the thesis aims to improve the accuracy of variational motion estimation methods. The research intends to solve common problems in motion estimation such as motion discontinuities, low-textured regions and occlusions by their efficient characterization. To this end, a novel sparse variational regularization technique was developed for variational methods. First, to evaluate the performance of the proposed regularization under noise and from unreliable measurements, it was applied to the problems of image denoising and image reconstruction from highly incomplete measurements. Second, utilizing this novel regularization, a method was proposed to estimate optical flow from noisy and unreliable image intensity measurements. The results demonstrate that a reliable optical flow can be obtained from highly incomplete measurements without much decrease in the accuracy. This is in contrast to the existing variational methods which use all of the measurements for the estimation of a dense optical flow. Third, the proposed method was extended to employ sparse variational regularization for the detection of sparse occlusion using optical flow in video sequences.

Extensive experimental results were presented to show the effectiveness of methods proposed in this work. This chapter summarizes main contributions of the thesis in Section 6.1, and suggests some improvements and future research directions in Section 6.2.

6.1 Summary

This thesis explores the methods to estimate visual motion by mapping the co-visible scene points and occlusion detection across video frames. We have estimated optical flow between two consecutive video frames for the movement of visible pixels. Commonly encountered problems of motion discontinuities, occlusions, low-textured regions and illumination change are addressed by using variational methods. These methods are chosen mainly because of their ability to integrate application specific information in a consistent mathematical framework. The availability of efficient energy minimization methods is also a key factor for their selection.

Chapter 2 reviews literature on optical flow estimation methods. Variational methods are focussed because proposed methods utilize variational framework. Existing methods are discussed according to their capabilities to solve problems inherent to optical flow estimation. Different techniques are gradually introduced to the reader which can be combined together to improve the estimation quality of optical flow.

Chapter 3 presents a novel sparse variational regularizer HVD which is based on robust ℓ_1 norm. The proposed regularizer is able to preserve sharp horizontal, vertical and diagonal edges. This regularizer has been shown to reduce undesirable staircase artefacts in low-textured regions as produced by other edge preserving regularizers. The HVD regularizer has been applied to problems of image denoising and image restoration from incomplete measurements. Experimental results presented in this chapter show the superiority of the proposed regularizer over existing TV based regularizers.

Chapter 4 employs the proposed sparsity enhancing variational regularizer of Chapter 3 for optical flow estimation. Motion discontinuities are problematic for optical flow methods. To improve the quality of the estimated optical flow, the preservation of the

boundaries of the moving objects is essential. The HVD regularizer addresses this issue by exploiting the sparsity of the motion vectors in the gradient domain. Since HVD is convex, the energy of the regularization term is also convex. A first order, fast and efficient algorithm, which is suitable for large-scale convex optimization problems, has also been presented to minimize the variational energy. The efficacy of the proposed method to estimate optical flow is demonstrated by extensive experimental results. Our experiments show that optical flow can be estimated from as low as 10% of measurements by exploiting the sparsity of partial flow derivatives. Thus, a decrease in the computational complexity and memory requirements is achieved.

Chapter 5 solves the problem of joint optical flow estimation and sparse occlusion detection under changing lighting conditions of the scene. Illumination change is one of the main problems in motion estimation. The mapping of pixels may be wrong or not possible at all under varying illumination. If the brightness change is not handled properly, methods for occlusion detection using optical flow can incorrectly detect occlusion in those co-visible video regions where the brightness of the scene points is changing. To the contrary, under brightness change scenarios, these methods may fail to detect true occluded regions. Since occluded regions are generally small in any frame of a video sequence, the proposed method presented in this chapter imposes a sparsity constraint on these occluded regions. Moreover, the optical flow constraint is replaced by a generalized dynamic image model (GDIM) to form the pixel matching data term. The brightness change is thus captured in parameters of the GDIM, and these parameters are also estimated along with optical flow and occlusion. A computationally less expensive version of the proposed method is also presented which does not require the estimation of brightness change parameters under certain conditions. Experiments on both synthetic and real video sequences have demonstrated that the proposed method show robustness to brightness constancy violations while detecting the occluded video regions.

6.2 Future Directions

Though methods proposed in this work are able to obtain promising optical flow estimation and occlusion detection results by utilizing concepts from sparse signal techniques, there is still room for the improvement. The research in this thesis can be broadened in the following ways.

- The proposed HVD regularizer minimizes the ℓ_1 norm of derivatives of an unknown image to be restored. It promotes more sparsity than total variation regularizer by enforcing the continuity of partial image derivatives at each pixel, and minimizing their ℓ_1 norm separately. The HVD regularizer also avoids undesirable staircase artefacts as produced by total variation; therefore, it can be applied to other vision problems which use TV, and suffer from shortcomings of TV. Some examples include super resolution, three-dimensional motion estimation, structure from motion and multi-view scene reconstruction.

- A linear convex method, presented in [150], has been modified and then used for problems addressed in Chapter 3, 4 and 5. Numerous other methods, for example, split Bregman method [171], gradient projection for sparse reconstruction (GPSR) [149], and Bayesian compressive sensing [172] have been proposed for the recovery of sparse signals. These methods have not been used in this thesis. Their performances for problems dealt in this thesis are yet to be explored.

- Several strategies have been used for the measurement of intensity derivatives in Chapter 4. It is found that a combination of significant and random measurements work well. Significant derivative measurements, which lie on intensity edges, are needed to preserve motion discontinuities because usually image and motion boundaries coincide in a video frame. Random measurements are also needed especially in low-textured video regions so that the algorithm does not trap there due to insufficient measurements. Since the proposed method uses a coarse-to-fine strategy to handle large displacements, details of tiny objects may get lost at coarser levels

of the pyramid. Special care has to be taken for measurements related to these objects. Measurements can be made adaptive to the size of small objects or some sophisticated technique, for example, an extended coarse-to-fine refinement framework can be used to address this issue [82]. This framework reduces the trust of coarser level optical flow estimates propagated to finer levels of the pyramid; thus, it enables the recovery of motion details at each level of the pyramid.

- The proposed method uses cues from the residual of the brightness matching term to estimate optical flow in Chapter 5. This residual is usually small when the motion of moving objects is small. Our experiments show that the proposed method can successfully detect occlusion when a sufficient interframe motion of objects is present in the scene. However, the occlusion detection precision of this method decreases in case of very small motion of objects. Static occlusion cues, for example, brightness, color and texture have been found to work better in this case. These static image cues have not been embedded into the proposed method for the detection of occluded video regions. The integration of these cues with dynamic cues is found to enhance occlusion detection capabilities of optical flow estimation methods [170]. The learning of occlusion boundaries based on static cues has also been found to be useful [173]. It remains to be explored that how machine learning techniques and static image cues can be combined with dynamic cues to improve the performance of the proposed sparse occlusion detection method. Some occlusion detection methods use a number of video frames to detect occlusion. The accuracy of the proposed method can be increased by using static and dynamic occlusion cues over multiple video frames.

- Parameters to weight the occlusion and regularization terms have been selected by hit and trial in the sparse occlusion detection method proposed in Chapter 5. Such a practice is rather cumbersome; thus, optical flow and occlusion detection results are affected by the choice of these parameters. Parameters to control the tradeoff between the sparse occlusion, the regularization and data fidelity terms can

be chosen adaptively. The method, presented in [174], utilizes a Bayesian framework to estimate regularization parameters in optical flow estimation. This method can be useful in the proposed sparse occlusion detection method for the estimation of regularization parameters.

- This thesis does not consider the implementation of proposed methods on the hardware. Bidirectional multi-grid methods have been successfully used for the hardware implementation of variational optical flow methods in real time [175]. It is still to be researched that how proposed methods can get benefit from such hardware implementations.

Bibliography

[1] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2010.

[2] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, "An improved algorithm for tv-l1 optical flow," in *Statistical and Geometrical Approaches to Visual Motion Analysis* (D. Cremers, B. Rosenhahn, A. L. Yuille, and F. R. Schmidt, eds.), pp. 23–45, Springer-Verlag, 2009.

[3] A. Wedel, D. Cremers, T. Pock, and H. Bischof, "Structure and motion adaptive regularization for high accuracy optic flow," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1663–1668, Sep. 29-Oct. 2, 2009.

[4] D. Sun, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2432–2439, June 13-18, 2010.

[5] A. Ayvaci, M. Raptis, and S. Soatto, "Sparse occlusion detection with optical flow," *International Journal of Computer Vision*, vol. 97, no. 3, pp. 322–338, 2012.

[6] A. Stein and M. Hebert, "Occlusion boundaries from motion: Low-level detection and mid-level reasoning," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 325–357, 2009.

[7] A. Verri and T. Poggio, "Motion field and optical flow: qualitative properties," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 5, pp. 490–498, 1989.

[8] G. Lefaix, E. Marchand, and P. Bouthemy, "Motion-based obstacle detection and tracking for car driving assistance," in *Proceedings of the IEEE International Conference on Pattern Recognition*, pp. 74 – 77, Aug. 11-15, 2002.

[9] G. N. Desouza and A. C. Kak, "Vision for mobile robot navigation: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 237–267, 2002.

[10] D. DeCarlo and D. N. Metaxas, "Optical flow constraints on deformable models with applications to face tracking," *International Journal of Computer Vision*, vol. 38, no. 2, pp. 99–127, 2000.

[11] H. Mallot, H. H. Bülthoff, J. J. Little, and S. Bohrer, "Inverse perspective mapping simplifies optical flow computation and obstacle detection," *Biological Cybernetics*, vol. 64, no. 3, pp. 177–185, 1991.

[12] P. Bouttefroy, *Contextual Bayesian inference for visual object tracking and abnormal behaviour detection*. PhD thesis, School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, 2010.

[13] X. Wu, Y. Ou, H. Qian, and Y. Xu, "A detection system for human abnormal behavior," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1204–1208, Aug. 2-6, 2005.

[14] S. C. Han and C. I. Podilchuk, "Video compression with dense motion fields," *IEEE Transactions on Image Processing*, vol. 10, no. 11, pp. 1605–1612, 2001.

[15] H.-C. Chang, S.-H. Lai, and K.-R. Lu, "A robust and efficient video stabilization algorithm," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, vol. 1, pp. 29–32, June 27-30, 2004.

[16] J. Klappstein, T. Vaudreyi, C. Rabe, A. Wedel, and R. Klette, "Moving object segmentation using optical flow and depth information," in *Proceedings of the 3rd Pacific Rim Symposium on Advances in Image and Video Technology*, pp. 611–623, Jan. 13-16, 2008.

[17] G. Adiv, "Determining three-dimensional motion and structure from optical flow generated by several moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 4, pp. 384–401, 1985.

[18] R. Szeliski, "Image alignment and stitching: a tutorial," *Foundations and Trends in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2006.

[19] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *Proceedings of the European Conference on Computer Vision*, pp. 428–441, May 7-13, 2006.

[20] R. Chaudhry, A. Ravichandran, G. D. Hager, and R. Vidal, "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1932–1939, June 20-26, 2009.

[21] Y. Yacoob and L. S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 636–642, 1996.

[22] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 8, pp. 185–203, 1981.

[23] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, vol. 2, no. 3, pp. 283–310, 1989.

[24] D. Marr and S. Ullman, "Directional selectivity and its use in early visual processing," *Proceedings of the Royal Society of London, Series B, Biological Sciences*, vol. 211, no. 1183, pp. 151–180, 1981.

[25] J. J. Little, H. H. Bülthoff, and T. Poggio, "Parallel optical flow using local voting," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 454–459, Jan. 4-7, 1988.

[26] M. A. Sutton, W. J. Wolters, W. H. Peters, W. F. Ranson, and S. R. McNeill, "Determination of displacements using an improved digital correlation method," *Image and Vision Computing*, vol. 1, no. 3, pp. 133Ű–139, 1983.

[27] G. Bonmassar and E. L. Schwartz, "Improved cross-correlation for template matching on the laplacian pyramid," *Pattern Recognition Letters*, vol. 19, no. 8, pp. 765 – 770, 1998.

[28] J. van de Weijer and T. Gevers, "Robust optical flow from photometric invariants," in *Proceedings of the IEEE international conference on image processing*, p. 1835Ű1838, Oct. 24-27, 2004.

[29] A. Bruhn and J. Weickert, "Towards ultimate motion estimation: combining highest accuracy with real-time performance," in *Proceedings of the IEEE international conference on computer vision*, p. 749Ű755, Oct. 17-20 2005.

[30] H. Zimmer, A. Bruhn, and J. Weickert, "Optic flow in harmony," *International Journal of Computer Vision*, vol. 93, no. 3, pp. 368–388, 2011.

[31] P. Golland and A. M. Bruckstein, "Motion from color," *Computer Vision and Image Understanding*, vol. 68, no. 3, p. 346Ű362, 1997.

[32] F. Zhao, Q. Huang, and W. Gao, "Image matching by normalized cross-correlation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 729–732, May 15-19, 2006.

[33] F. Zhao, Q. Huang, and W. Gao, "Image matching by normalized cross-correlation," in *In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2006.

[34] J. Lewis, "Fast template matching," *Vision Interface*, vol. 95, pp. 120–123, 1995.

[35] C. Sun, "Fast optical flow using cross correlation and shortest-path techniques," in *In Proceedings of the Digital Image Computing: Techniques and Applications*, pp. 143–148, 1999.

[36] S. Mattoccia, F. Tombari, and L. D. Stefano, "Reliable rejection of mismatching candidates for efficient zncc template matching," in *In Proceedings of the IEEE International Conference on Image Processing*, pp. 849–852, 2008.

[37] A. Mahmood and S. Khan, "Correlation-coefficient-based fast template matching through partial elimination.," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2099–2108, 2012.

[38] A. Singh, "An estimation-theoretic framework for image-flow computation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 168–177, Dec. 4-7, 1990.

[39] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the International Joint Conferences on Artificial Intelligence*, pp. 674Ű–679, Aug. 24-28, 1981.

[40] H. H. Nagel, "Displacement vectors derived from second-order intensity variations in image sequences," *Computer Vision, Graphics, and Image Processing*, vol. 21, no. 1, pp. 85–117, 1983.

[41] S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational appraoch to motion perception," *Biological Cybernetics*, vol. 60, pp. 79–97, 1988.

[42] S. Baker and I. Matthews, "Lucas-kanade 20 years on: a unifying framework: part 1," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.

[43] J. Y. Bouguet, "Pyramidal implementation of the lucas kanade feature tracker: description of the algorithm," tech. rep., Intel Corporation, Microprocessor Research Labs, 2000.

[44] A. Bab-Hadiashar and D. Suter, "Robust optic flow computation," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 59–77, 1998.

[45] J. Weber and J. Malik, "Robust computation of optic flow in a multiscale differential framework," *International Journal of Computer Vision*, vol. 14, no. 1, pp. 67–81, 1995.

[46] C. Tomasi and T. Kanade, "Detection and tracking of point features," tech. rep., Carnegie Mellon University, 1991.

[47] G. D. Hager and P. N. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 10, pp. 1025–1039, 1998.

[48] H. Jin, P. Favaro, and S. Soatto, "Real-time feature tracking and outlier rejection with changes in illumination," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 684 – 689, July 7-14, 2001.

[49] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593Ű–600, June 21-23, 1994.

[50] N. Dowson and R. Bowden, "Mutual information for lucas-kanade tracking (milk): An inverse compositional formulation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 1, pp. 180–185, 2008.

[51] E. P. Simoncelli, E. Adelson, and D. J. Heeger, "Probability distributions of optical flow," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 310–315, June 3-6, 1991.

[52] V. M. Govindu, "Revisiting the brightness constraint: probabilistic formulation and algorithms," in *Proceedings of the European Conference on Computer Vision*, pp. 177–188, May 7 - 13, 2006.

[53] Y. Weiss and D. J. Fleet, "Velocity likelihoods in biological and machine vision," in *Probabilistic Models of the Brain: Perception and Neural Function*, pp. 81–100, MIT Press, 2001.

[54] M. Okutomi and T. Kanade, "A locally adaptive window for signal matching," *International Journal of Computer Vision*, vol. 7, no. 2, pp. 143–162, 1992.

[55] M. Black and A. Jepson, "Estimating multiple independent motions in segmented images using parametric models with local deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 972–986, 1996.

[56] C. L. Zitnick, N. Jojic, and S. B. Kang, "Consistent segmentation for optical flow estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1308 – 1315, Oct. 17-20, 2005.

[57] X. Ren, "Local grouping for optical flow," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 24-26, 2008.

[58] J. Malik, S. Belongie, T. K. Leung, and J. Shi, "Contour and texture analysis for image segmentation," *International Journal of Computer Vision*, vol. 43, no. 1, pp. 7–27, 2001.

[59] M. J. Black and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[60] J. R. Bergen, P. Anandan, T. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proceedings of the European Conference on Computer Vision*, pp. 237–252, May 19-22, 1992.

[61] M. J. Black and A. Jepson, "Estimating optical flow in segmented images using variable-order parametric models with local deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 972–986, 1996.

[62] H. Sawhney and S. Ayer, "Compact representations of videos through dominant and multiple motion estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 814–830, 1996.

[63] J. M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of Visual Communication and Image Representation*, vol. 6, no. 4, pp. 348–365, 1995.

[64] J. Yang and H. Li, "Dense, accurate optical flow estimation with piecewise parametric model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1019–1027, 2015.

[65] M. Irani, P. Anandan, and M. Cohen, "Direct recovery of planar-parallax from multiple frames," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1528Ű–1534, 2002.

[66] A. Shashua and Y. Wexler, "Q-warping: Direct computation of quadratic reference surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 920Ű–925, 2001.

[67] H. yeung Shum and R. Szeliski, "Construction of panoramic image mosaics with global and local alignment," *International Journal of Computer Vision*, vol. 36, no. 2, pp. 101–130, 2000.

[68] R. Szeliski and J. Coughlan, "Spline-based image registration," *International Journal of Computer Vision*, vol. 22, no. 3, pp. 199–218, 1997.

[69] B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, and N. Navab, "Optical flow estimation with uncertainties through dynamic mrfs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1 – 8, June 24-26, 2008.

[70] R. Szeliski and H.-Y. Shum, "Motion estimation with quadtree splines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 12, pp. 1199Ű–1207, 1996.

[71] J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1427–1442, 2003.

[72] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.

[73] A. A. Amini, R. W. Curwen, and J. C. Gore, "Snakes and splines for tracking non-rigid heart motion," in *Proceedings of the European Conference on Computer Vision*, pp. 251–261, Apr. 15-18, 1996.

[74] N. B. Karayiannis, S. Member, Y. Xiong, J. D. F. Jr., M. S. Wise, and E. M. Mizrahi, "Quantifying motion in video recordings of neonatal seizures by regularized optical flow methods," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 890–903, 2005.

[75] S. Klein, M. Staring, and J. P. W. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2879Ű–2890, 2007.

[76] A. Bruhn, J. Weickert, and C. Schnorr, "Lucas/kanade meets horn/schunck: combining local and global optic flow methods," *International Journal of Computer Vision*, vol. 61, no. 3, pp. 211–231, 2005.

[77] J. Weickert and C. Schnorr, "Variational optic flow computation with a spatio-temporal smoothness constraint," *Journal of Mathematical Imaging and Vision*, vol. 14, no. 3, pp. 245–255, 2001.

[78] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J.Weickert., "Highly accurate optic flow computation with theoretically justified warping," *International Journal of Computer Vision*, vol. 67, no. 2, pp. 141–158, 2006.

[79] M. J. Black, *Robust Incremental Optical Flow.* PhD thesis, Department of Computer Science, Yale University, 1992.

[80] V. S. Lempitsky, S. Roth, and C. Rother, "Fusionflow: Discrete-continuous optimization for optical flow estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1 – 8, June 24-26, 2008.

[81] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof, "Anisotropic huber l1 optical flow," in *Proceedings of the British Machine Vision Conference*, pp. 1–11, Sep. 7-10, 2009.

[82] L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1744–1757, 2012.

[83] R. Timofte and L. V. Gool, "Sparse flow: Sparse matching for small to large displacement optical flow," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1100–1106, 2015.

[84] E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," *Journal of the Optical Society of America A*, vol. 2, no. 2, pp. 284–299, 1985.

[85] D. Heeger, "Optical flow using spatiotemporal filters," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 279–302, 1988.

[86] J. Biguan, G. Granlund, and J. Wiklund, "Multidimensional orientation estimation with application to texture analysis and optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 775–790, 1991.

[87] D. Fleet and A. Jepson, "Computation of component image velocity from local phase information," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 77–104, 1990.

[88] J. Weickert, A. Bruhn, T. Brox, and N. Papenberg, "A survey on variational optic flow methods for small displacements," in *Mathematical Models for Registration and Applications to Medical Imaging* (O. Scherzer, ed.), vol. 10, pp. 103–136, Springer, 2006.

[89] J. L. Barron, D. J. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[90] D. J. Fleet and Y. Weiss, "Optical flow estimation," in *Handbook of Mathematical Models in Computer Vision* (N. Paragios, Y. Chen, and O. Faugeras, eds.), pp. 239–257, Springer, 2005.

[91] T. Brox, A. Bruhn, N. Papenberg, and J.Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proceedings of the European Conference on Computer Vision*, pp. 25Ű–36, May 11-14, 2004.

[92] C. Zach, T. Pock, and H. Bischof, "A duality based approach for real time tv-l1 optical flow," in *Proceedings of the DAGM conference on Pattern recognition*, pp. 214–223, Sep. 12-14, 2007.

[93] D. Wei, C. Liu, and W. T. Freeman, "A data-driven regularization model for stereo and flow," in *Proceedings of the 2nd International Conference on 3D Vision (3DV)*, pp. 277–284, 2014.

[94] O. Tretiak and L. Pastor, "Velocity estimation from image sequences with second order differential opertors," in *Proceedings of the IEEE International Conference on Pattern Recognition*, pp. 16–19, 1984.

[95] C. Schnorr, "Segmentation of visual motion by minimizing convex non-quadratic functionals," in *Proceedings of the IEEE International Conference on Pattern Recognition*, pp. 661 – 663, Oct. 9-13, 1994.

[96] L. Chen, J. Wang, and Y. Wu, "Decomposing and regularizing sparse/non-sparse components for motion field estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1776–1783, June 16-21, 2012.

[97] H. W. Haussecker and D. J. Fleet, "Computing optical flow with physical models of brightness variation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 661–673, 2001.

[98] S. Negahdaripour, "Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 9, pp. 961–979, 1998.

[99] H. H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 5, pp. 565–593, 1986.

[100] L. Álvarez, R. Deriche, T. Papadopoulo, and J. Sánchez, "Symmetrical dense optical flow estimation with occlusions detection," *International Journal of Computer Vision*, vol. 75, no. 3, pp. 371–385, 2007.

[101] T. H. Kim, H. S. Lee, and K. M. Lee, "Optical flow via locally adaptive fusion of complementary data costs," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3344–3351, 2013.

[102] R. Deriche, P. Kornprobst, and G. Aubert, "Optical flow estimation while preserving its discontinuities: a variational approach," in *Proceedings of the Asian Conference on Computer Vision*, pp. 290–295, Dec. 5-8, 1995.

[103] J. Weickert and C. Schnorr, "A theoretical framework for convex regularizers in pde-based computation of image motion," *International Journal of Computer Vision*, vol. 45, no. 3, pp. 245–264, 2001.

[104] J. J. Gibson, *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, 1986.

[105] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister, "Stereo matching with color-weighted correlation, hierachical belief propagation and occlusion handling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 492–504, 2009.

[106] J. Sun, Y. Li, and S. B. Kang, "Symmetric stereo matching for occlusion handling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 399–406, June 20-26, 2005.

[107] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions via graph cuts," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 508–515, July 7-14, 2001.

[108] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993–1008, 2003.

[109] G. Egnal and R. P. Wildes, "Detecting binocular half-occlusions: Empirical comparisons of five approaches," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1127–1133, 2002.

[110] S. Ince and J. Konrad, "Occlusion-aware optical flow estimation," *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1443–1451, 2008.

[111] M. Proesmans, L. van Gool, E. Pauwels, and A. Oosterlinck, "Determination of optical flow and its discontinuities using non-linear diffusion," in *Proceedings of the European conference on Computer Vision*, pp. 295–304, May 2-6, 1994.

[112] J. Xiao, H. Cheng, H. Sawhney, C. Rao, and M. Isnardi, "Bilateral filtering-based optical flow estimation with occlusion detection," in *Proceedings of the European Conference on Computer Vision*, pp. 211–224, May 7-13, 2006.

[113] P. Sand and S. J. Teller, "Particle video: Long-range motion estimation using point trajectories," *International Journal of Computer Vision*, vol. 80, no. 1, pp. 72–91, 2008.

[114] E. Parrilla, D. Ginestar, J. L. Hueso, J. Riera, and J. R. Torregrosa, "Handling occlusion in optical flow algorithms for object tracking," *Computers & Mathematics with Applications*, vol. 56, no. 3, pp. 733 – 742, 2008.

[115] R. Kennedy and C. J. Taylor, "Optical flow with geometric occlusion estimation and fusion of multiple frames," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 364–377, Springer, 2015.

[116] K. Lim, A. Das, and M. Chong, "Estimation of occlusion and dense motion fields in a bidirectional bayesian framework," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 712–718, 2002.

[117] C. Strecha, R. Fransens, and L. J. V. Gool, "A probabilistic approach to large displacement optical flow and occlusion detection," in *Proceedings of the European Conference on Computer Vision Workshop on Statistical Methods in Video Processing*, pp. 71–82, May 16, 2004.

[118] D. Sun, E. B. Sudderth, and M. J. Black, "Layered image motion with explicit occlusions, temporal consistency, and depth ordering," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 2226–2234, Dec. 6-9, 2010.

[119] D. Sun, C. Liu, and H. Pfister, "Local layering for joint motion estimation and occlusion detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1098–1105, 2014.

[120] I. M. Gelfand and S. V. Fomin, *Calculus of Variations.* Courier Dover Publications, 2000.

[121] D. L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind," *Journal of the ACM*, vol. 9, no. 1, pp. 84–97, 1962.

[122] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[123] R. H. Chan, T. Chan, and H. M. Zhou, "Continuition method for total variation denoising problems," Tech. Rep. 95-18, University of California, Los Angeles, 1995.

[124] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[125] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.

[126] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 460–489, 2005.

[127] T. Chan, G. Golub, and P. Mulet, "A nonlinear primal dual method for total variation-based image restoration," *SIAM Journal on Scientific Computing*, vol. 20, no. 6, pp. 1964–1977, 1999.

[128] M. A. Kitchener, A. Bouzerdoum, and S. L. Phung, "Adaptive regularization for image restoration using a variational inequality approach," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 2513–2516, Sep. 26-29 2010.

[129] T. F. Chan, A. M. Yip, and F. E. Park, "Simultaneous total variation image inpainting and blind deconvolution," *International Journal of Imaging Systems and Technology*, vol. 15, no. 1, pp. 92–102, 2005.

[130] T. F. Chan and K. C. Wong, "Total variation blind deconvolution," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.

[131] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. T.Figueiredo, "Adaptive total variation image deblurring: A majorization-minimization approach," *Signal Processing*, vol. 89, no. 9, pp. 1683–1693, 2009.

[132] M. W. Nawaz, A. Bouzerdoum, and S. L. Phung, "Motion estimation with adaptive regularization and neighborhood dependent constaint," in *Proceedings of the Digital Image Computing, Techniques and Analysis*, pp. 387 – 392, Dec. 1-3, 2010.

[133] X. Shen and Y. Wu, "Sparsity model for robust optical flow estimation at motion discontinuities," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2456 – 2463, June 13-18, 2010.

[134] M. Werlberger, T. Pock, and H. Bischof, "Motion estimation with non-local total variation regularization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2464–2471, June 13-18, 2010.

[135] N. Slesareva, A. Bruhn, and J. Weickert, "Optic flow goes stereo: A variational method for estimating discontinuity preserving dense disparity maps," in *Proceedings of the27th DAGM Symposium*, pp. 33–40, Aug 31 - Sep. 2 2005.

[136] R. Ben-Ari and N. A. Sochen, "Variational stereo vision with sharp discontinuities and occlusion handling," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1–7, Oct. 14-20 2007.

[137] W. Miled, J.-C. Pesquet, and M. Parent, "Disparity map estimation using a total variation bound.," in *Proceedings of the Third Canadian Conference on Computer and Robot Vision*, pp. 48–55, June 7-9 2006.

[138] S. Esedoglu and S. J. Osher, "Decomposition of images by the anisotropic rudin-osher-fatemi model," *Communications on Pure and Applied Mathematics*, vol. 57, no. 1, pp. 1609–1626, 2004.

[139] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[140] L. Wanga, L. Xiaoa, J. Zhangb, and Z. Weia, "New image restoration method associated with tetrolets shrinkage and weighted anisotropic total variation," *Signal Processing*, vol. 93, no. 4, p. 661Ű670, 2013.

[141] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling & Simulation*, vol. 7, no. 3, pp. 1005–1028, 2008.

[142] G. Peyre, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *Proceedings of the European Conference on Computer Vision* (D. Forsyth, P. Torr, and A. Zisserman, eds.), vol. 5304 of *Lecture Notes in Computer Science*, pp. 57–68, Springer Berlin Heidelberg, 2008.

[143] T. Chan and A. M. P. Mulet, "High-order total variation-based image restoration," *SIAM Journal on Scientific Computing*, vol. 22, no. 2, pp. 503–516, 2000.

[144] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.

[145] J. Weickert, *Anisotropic Diffusion in Image Processing.* B. G. Teubner, 1998.

[146] D. Goldfarb and W. Yin, "Parametric maximum flow algorithms for fast total variation minimization," *SIAM Journal on Scientific Computing*, vol. 31, no. 5, pp. 3712–3743, 2009.

[147] A. Chambolle., "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging*, vol. 20, no. 1, pp. 89–97, 2004.

[148] Y. Nesterov, "Gradient methods for minimizing composite objective function," tech. rep., Center for operations research and econometrics, Catholic University of Louvein, 2007.

[149] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.

[150] S. Becker, J. Bobin, and E. J. Candes, "Nesta: a fast and accurate first-order method for sparse recovery," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 1–39, 2011.

[151] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[152] J. Bioucas-Dias and M. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2992 – 3004, 2007.

[153] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[154] A. Chambolle, "Total variation minimization and a class of binary mrf models," in *Energy Minimization Methods in Computer Vision and Pattern Recognition* (S. chun Zhu, A. L. Yuille, and D. Cremers, eds.), pp. 136–152, Springer Berlin / Heidelberg, 2005.

[155] J. Darbon and M. Sigelle, "Image restoration with discrete constrained total variation, part i: fast and exact optimization," *Journal of Mathematical Imaging and Vision*, vol. 26, no. 3, p. 261Ű276, 2006.

[156] R. Ranftl, K. Bredies, and T. Pock, "Non-local total generalized variation for optical flow estimation," in *Proceedings of the European Conference on Computer Vision*, pp. 439–454, 2014.

[157] X. Shu and N. Ahuja, "Hybrid compressive sampling via a new total variation tvl1," in *Proceedings of the European Conference on Computer Vision*, pp. 393–404, Sep. 5 -11, 2010.

[158] K. Choi, J. Wang, L. Zhu, T.-S. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method," *Medical Physics*, vol. 37, no. 9, pp. 5113–5125, 2010.

[159] J. Dahl, P. C. Hansen, S. H. Jenson, and T. L. Jensen, "Algorithms and software for total variation image reconstruction via first-order methods," *Numerical Algorithms*, vol. 53, no. 1, pp. 67–92, 2009.

[160] A. N. Tikhonov, "On the stability of inverse problems," *Doklady Akademii Nauk SSSR*, vol. 39, no. 5, p. 195Ű198, 1943.

[161] S. Roth and M. J. Black, "On the spatial statistics of optical flow," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 33–50, 2007.

[162] D. sun, J. P. Lewis, and M. Black, "Learning optical flow," in *Proceedings of the European Conference on Computer Vision*, pp. 83–97, Oct. 12-18, 2008.

[163] M. W. Nawaz, A. Bouzerdoum, and S. L. Phung, "Optical flow estimation using sparse gradient representation," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 2681 – 2684, Sep. 11-14, 2011.

[164] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.

[165] C. Liu, W. Freeman, E. Edelson, and Y. Weiss, "Human-assisted motion annotation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 24-26, 2008.

[166] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[167] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[168] Y.-H. Kim, A. M. Martínez, and A. C. Kak, "Robust motion estimation under varying illumination," *Image and Vision Computing*, vol. 23, no. 4, pp. 365–375, 2005.

[169] S.-H. Lai, "Robust image matching under partial occlusion and spatially varying illumination change," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 84–98, 2000.

[170] P. Sundberg, T. Brox, M. Maire, P. Arbelaez, and J. Malik, "Occlusion boundary detection and figure/ground assignment from optical flow," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2233–2240, June 20-25, 2011.

[171] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[172] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[173] D. Martin, C. Fowlkes, and J. Malik., "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 530Ű–549, 2004.

[174] P. Héas, C. Herzet, and É. Mémin, "Bayesian inference of models and hyperparameters for robust optical-flow estimation," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1437–1451, 2012.

[175] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnorr, "Variational optical flow computation in real time," *IEEE Transactions on Image Processing*, vol. 14, no. 5, pp. 608–615, 2005.