ROBUST MOTION ESTIMATION AND ANALYSIS BASED ON STATISTICAL INFORMATION

Thesis submitted in partial fulfillment
of the requirements for the degree of

MS by Research
in
Electronics and Communications

by

V.S.RAO VEERAVASARAPU
200731010
vsrao.v@research.iiit.ac.in

CENTRE FOR VISUAL INFORMATION TECHNOLOGY
International Institute of Information Technology
Hyderabad - 500 032, INDIA
February 2013

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled *Robust Motion Estimation and Analysis based on Statistical Information* by V.S.Rao Veeravasarapu, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date

Adviser: Prof. Jayanthi Sivaswamy

To my Family

Acknowledgments

I owe a great intellectual debt to my advisor Prof. Jayanthi Sivaswamy who gave me the introduction to signal and image processing and vision research fields and who has an uncanny ability to find the weak link in a chain of reasoning. Her encouragement and guidance made this work possible. The knowledge and valuable suggestions given by her provided the foundation for the work presented in this thesis.

I am also grateful for the constant support and encouragement of Gopal Datt Joshi, my secondary guide and friend who gave me valuable feedback, suggestions and inputs. I should be thankful to him for the cooperation in writing of my first paper which is accepted at ICVGIP-10. I also want to thank the gang at MIP lab for long walk discussions and particularly N.V.Kartheek for his support of all things technical. I must also thank Harini, who taught me about how to understand a research paper, for providing the chance to work speech processing related areas, from which this work got benefits. This thesis has also benefited from discussions in group meetings, as well as from comments by Sai deepak, Sandeep kaushik and Mayank chawla. My discussions with Akhilesh, Shashank, and Rohit has been both productive and fun. I would like to thank Vidya who collected the ultrasound cardiac data and the people who are maintaining the Middlebury optical flow evaluation on-line portal for providing the data for optical flow estimation.

I would like to especially thank Prof. P.J. Narayanan, Jawahar and Anoop for providing a wonderful research environment at CVIT. Last and but not least, I thank Dr. Madhava Krishna and Dr. Kishore Prahallad for giving valuable feedback in reviewing process.

Abstract

Accurate and Robust estimation of optical flow continues to be of interest due to the deep penetration of digital cameras into many areas including robot navigation and video surveillance applications. The canonical approach to the flow estimation relies on local brightness constancy which has limitations. In this thesis, we re-examine the optic flow problem and formulate an alternate hypothesis that optical flow is an apparent motion of local information across frames and propose a novel framework to robustly estimate flow parameters. Pixel-level matching approach has been implemented according to the proposed formulation in which optical flow is estimated based on local information associated with each pixel. Self information and a variety of divergence measures have been investigated for capturing the local information. Results of benchmarking with the Middlebury dataset show that the proposed formulation is comparable to the top performing methods in accurate flow computation. The distinguishing aspects however are that these results hold for small as well as large displacements and the flow estimation is robust to distortions such as noise, illumination changes, non-uniform blur etc. Thus, the local information based approach offers a promising alternative to computing optical flow. We also developed a method to remove motion blur from frames by using the information measures. The effectiveness of the proposed motion estimation approach is also demonstrated on extraction of structure from motion of synthetic microtexture patterns, cardiac ultrasound sequences and colorization of black and white videos.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Visual motion is the 2D velocity field corresponding to the movement of brightness patterns in the image plane of a camera/eye. In general, it depends on the relative motion between 3D objects and the camera, and provides rich information about the structures of the objects and their dynamic behaviour. Human beings rely on the skills of perceiving and understanding visual motion in order to move around, meet with people, watch movies and perform many other essential daily tasks. If we want computers to assist us and interact with us, we must endow them with a similar capability for analyzing visual motion, that is, accurately measuring and appropriately interpreting the 2D motion present in digital images. This has turned out to be a highly complicated and error-prone process. The co-existence of profound significance and great challenge makes visual motion analysis a very important and active research area in computer vision.

Optical flow is a flexible representation of visual motion that is particularly suitable for computers analyzing digital images. It associates each image pixel with a flow vector (v_x, v_y) or (v, θ), indicating apparent instantaneous 2-D velocity. Optical flow (OF) computation is a classical problem in vision. It underlies robust and accurate motion estimation in key applications such as automatic robot navigation, tracking objects/people, derivation of structure from motion from noisy data such as cardiac ultrasound, etc. One of the first important studies on the computation of optical flow was done by Horn and Schunk [35] in the year 1981. According to their work, OF is defined as follows.

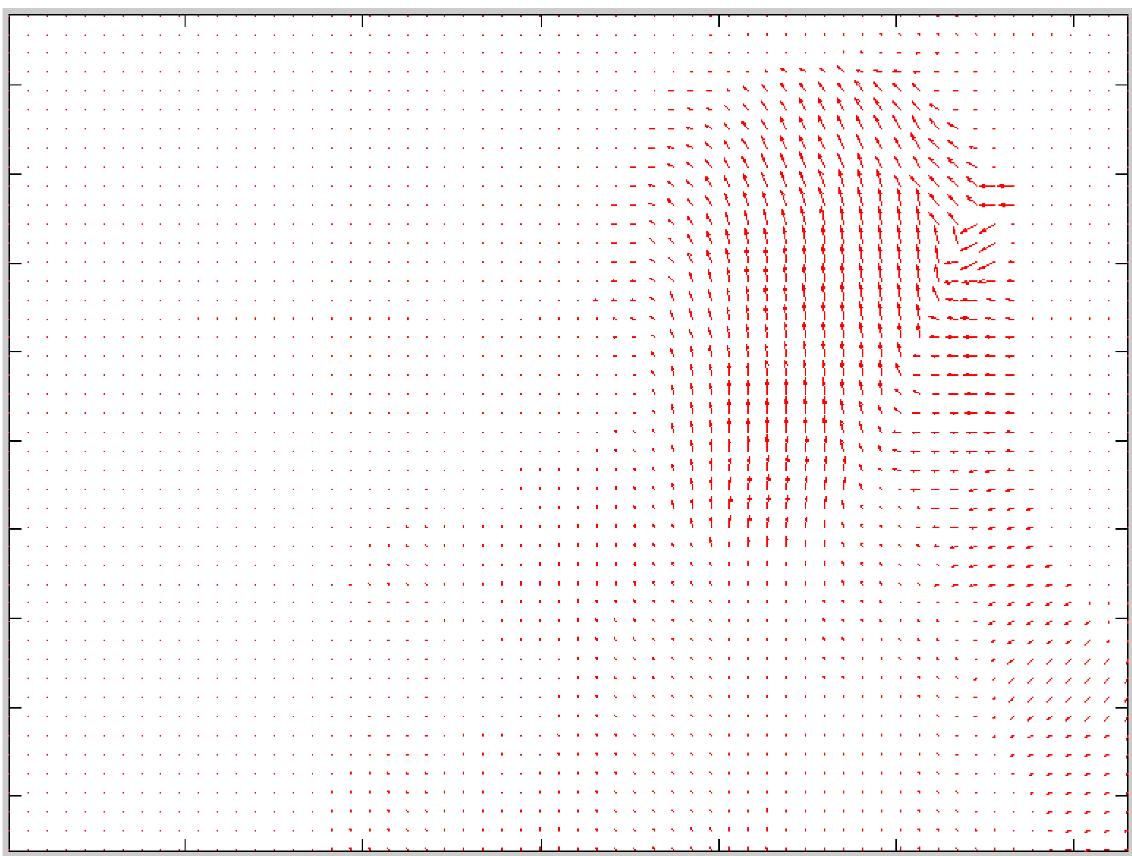
The optical flow is a velocity field in the image, which transforms one image into the next image in a sequence. As such it is not uniquely determined; the motion field, on the other hand, is a purely geometric concept, without any ambiguity - it is the projection into the image of three-dimensional motion vectors [36].

In order to illustrate the concept of OF, Fig.1.1 shows three frames that are part of a video sequence taken by a camera on a computer. The optical flow estimated for the tenth frame, which is subsampled by 10 to avoid being too crowded, is shown in Fig.1. It overall agrees with our perception of motion in the scene.

(a) frame09

(b) frame10

(c) frame11



(d) OF estimated between frame10 and frame11

Figure 1.1 Three frames in a video sequence taken by a camera on computer and the estimated OF field

1.1 Problem statement & Motivation

The canonical approach to estimating OF is based on matching patches using a brightness constancy (BC) assumption. However, this constraint is quite limited and is often invalid in real scenes, leading to difficulty in OF computation. In image regions that are roughly homogeneous, for instance, the OF is ambiguous because the BC assumption is satisfied by many different motions. This assumption is also violated by occlusions and by changing illumination, non-rigid motion, shadows, transparency, reflection etc. Additional constraints such as gradient constancy or regularization in an energy minimisation framework have been attempted. However, situations where these fail appear quite frequently in practice. For instance, gradient based approaches do not work well for low temporal sampling found in sequences obtained at low frames/sec, unexpected fast motions, etc. Hence, the problem of OF computation needs a fresh examination.

An alternative approach is to consider a pixel as a random variable (RV) and use local statistics for matching across frames. Even if the brightness of one pixel undergoes large changes between two consecutive frames, the local statistics of that pixel will be more stable and experience much less significant changes. Although, there have been several theories to measure information in a random variable, it has not been studied well in the context of motion analysis. Statistical information on the other hand, has been utilised in several ways in image alignment for medical image registration.

Two classes of measures are relevant for quantifying information: (i) Self information measures which relate to the information present in the single RV (ex: Shannon's entropy, Burg's entropy etc.); (ii) Mutual information based measures which relate to the dependence between two RVs (ex. Mutual Information, Kolmogorov divergence etc.). A vast number of entropy measures have been introduced in the Information theory literature generalizing Shannon's entropy. These are parametric, trigonometric and weighted entropies.

1.2 Contribution

In this work, we consider the statistical information such as divergence between the probability densities of RVs for OF estimation. Our motivation is to address the limitations in the BC-based OF formulation in vogue and explore an approach that does not trade off robustness for accuracy. The contribution of this work includes the following three aspects.

1. a new formulation for OF estimation in terms of a *local information* preservation rather than brightness preservation. This may provide a unified approach for handling small and large displacements while maintaining robustness to several degradations;

2. a pixel level matching method for robust as well as accurate OF estimation;

3. an extensive analysis of the impact of different information measures based on different divergence theories on OF computation;

1.3 Applications

The estimation of robust optical flow can be useful for many applications in computer vision and medical image processing etc. Hence, we also examined some selected applications from the above mentioned research fields where we need robust as well as accurate flow. Those are listed below.

- *Structure from Motion (Motion Edge Detection)* : We use the direction map of computed flow field to detect the edges formed by two different coherently moving patterns in different directions.
- *Heart-beat Rate estimation in Ultra-sonography* : We use the cyclic nature of flow vectors to estimate Heart rate from a given ultrasound cardiac image sequence.
- *Color Flow in Film Colorization* : We use flow fields to propagate color from most informative key frame (i.e. the frame which contains maximum number of frames).

We also extended the idea of *Statistical information optimization* to remove the effect of motion blur which is a common distortion in video capturing.

1.4 Organization

The first half of the thesis is devoted to optical flow estimation, and the rest describes the applications where we need robust as well as accurate flow. To enhance visual motion analysis robustness, which is the central issue in our study, statistical tools are extensively explored at every stage. Given the diversity of the topics, previous work is summarized and mathematical and statistical tools are introduced when the need arises.

Chapter 2: The common formulations of the optical flow problem are reviewed. The chapter first introduces the standard formulation of the brightness constancy constraint and then reviews three common techniques for flow estimation: gradient, block matching, and frequency based methods. In addition to describing the approaches, the chapter explores where they are violated and examines the current approaches for coping with motion discontinuities. Finally, we also discusses about the top ranked and well-known OF methods at Middlebury evaluation web site.

Chapter 3: The chapter introduces the robust estimation framework and uses it to reformulate the local and global approaches. We then test the effectiveness of frame work with different variants of divergence and discuss about some key implementation issues.

Chapter 4: The chapter discusses about the experiments we conducted to test the robustness and accuracy levels of the proposed formulation. As a part of this, detailed descriptions of the dataset and experimental results on real and synthetic images are presented. A survey on computational speeds are also presented.

Chapter 5: The chapter demonstrates the performance of the proposed formulation on three application where one needs robust as well as accurate flow fields. Those applications include (i) Motion edge detection: an algorithm to locate motion edge by using OF fields, (ii) Heartbeat rate estimation: a method to estimate heart rate from a given ultrasound cardiac image sequence by using cyclic movement of myocardial points, and (iii) Color flow: a method to propagate the colors from a automatically selected key frame by using flow vectors. For each application, experimental results on real and synthetic images are presented.

Chapter 6: An Information-based approach can also be used to address a common problem in capturing videos which is *motion blur*. This occurs due to the relative motion between the camera and the scene during the integration time. The chapter proposes a method to remove motion blur from the frames by using information measures with some additional constraints.

Chapter 7: We conclude by examining what questions have been answered and what questions remain open. In doing so we point to a number of future directions for work in optical flow.

Chapter 2

RELATED WORK

In this chapter, we present an overview on different techniques which are used to estimate the Optical Flow in conventional optical flow algorithms. Furthermore, we give examples of algorithms that implement the different techniques. For a comprehensive evaluation of optical flow algorithms we recommend the articles of Barron et al. [6] and McCane et al. [57].

Existing algorithms for OF determination can be divided into three broad categories:

- gradient based computation
- phase correlation based methods
- block matching methods

Apart from their differences, these techniques implement three processing stages. According to Barron et al. [6] these steps are as follows.

1. Noise reduction by applying low-pass or band-pass filters on the input frames.
2. Extraction of basic measurements, such as spatio-temporal derivatives (to measure normal components of velocity) or local correlation surfaces.
3. Integration of these measurements, to derive a two-dimensional motion field, which often involves assumptions on the motion field such as the motion varies smoothly.

We begin with the Brightness Constancy (BC) assumption as most of the OF methods depend on this assumption.

2.1 Brightness constancy assumption

Let $I(x, y, t)$ be the image intensity at a point (x, y) at time t. The BC assumption can be expressed as

$$I(x, y, t) = I(x + \partial x, y + \partial y, t + \partial t) \tag{2.1}$$

$$I(x, y, t) = I(x + v_x \partial t, y + v_y \partial t, t + \partial t) \tag{2.2}$$

where $(\partial x, \partial y)$ is the spatial displacement during the time interval ∂t, and (v_x, v_y) = $(\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t})$ is the optical flow vector. This equation simply states that a pixel maintains its intensity value during motion, or corresponding points in different frames have the brightness.

2.2 Gradient based methods

In gradient based techniques [35, 50, 15, 14], flow estimation depends on spatio-temporal derivatives of image intensity. These methods use approximations of the BC constraint Eq.2.2 by using Taylor series expansion.

$$I(x, y, t) = I(x, y, t) + I_x v_x \partial t + I_y v_y \partial t + I_t \partial t + \epsilon \tag{2.3}$$

where (I_x, I_y, I_t) is the image intensity gradient vector at the point (x, y, t), and ϵ represents the higher-order terms. If ϵ is negligible, the equation simplifies to

$$I_x v_x + I_y v_y + I_t = 0 \tag{2.4}$$

which is a linear equation with two unknowns v_x and v_y. This is called as the aperture problem of OF.

Given $n \geq 2$ pixels of the same motion, they can be grouped together and then v_x, v_y can be calculated through linear regression. Another way of obtaining constraints is to exploit second-order image derivatives. Differentiating Eq.2.4 with respect to x, y and t respectively gives three more equations:

$$I_{xx} v_x + I_{yx} v_y + I_{tx} = 0 \tag{2.5}$$

$$I_{xy} v_x + I_{yy} v_y + I_{ty} = 0 \tag{2.6}$$

$$I_{xt} v_x + I_{yt} v_y + I_{tt} = 0 \tag{2.7}$$

They can be used alone [6] or combined with Eq.2.4 [30] to solve for (v_x, v_y).

The most distinct attraction of gradient-based constraints, compared with matching-based constraints, is their ease of computation. The use of derivatives allows more efficient exploration of the solution space and hence achieves lower complexity and higher floating point precision [6]. However, it is important to point out that such advantages do come with a cost: the additional assumptions made in deriving the gradient-based constraints dictates a limited applicability. First of all, gradient-based constraints are valid only for small displacements, which in practice means magnitude <1 or 2 pixels/frame. Secondly, in order for the higher-order terms to be negligible, the local image intensity function should be close to a planar structure, which is also often violated. Finally, derivative estimation is a problematic process itself. It requires a linearisation step in order to make the computation tractable. This requires a warping strategy to handle large displacements, which can be problematic [87].

2.3 Block matching methods

Block matching methods [3, 48] find the flow vector or displacement that yields the best match between image regions in different frames. Best match can be defined in terms of maximizing a similarity measure such as the normalized cross-correlation (NCC), or minimizing a distance measure such as the sum-of-absolute-differences (SAD) and sum-of-squared-difference (SSD):

$$(v_x, v_y) = argmin_{\mathbf{v}} \sum_{(x,y)\epsilon B} [I(x, y, t) - I(x + v_x\partial t, y + v_y\partial t, t + \partial t)]^2 \tag{2.8}$$

where B is the set of pixels spanned by the block.

The block matching strategy provides a good trade-off between complexity and efficiency [48]. Several block matching criteria have been proposed using the SSD and SAD between two blocks. The popularity of these two criteria is due to their computational simplicity and performance. However, they do not consider the nature of the image and the information contained in a frame. These methods are severely limited in handling any image degradation. A variant that addresses this problem to some extent is an entropy based block matching [11]. This uses a k-th nearest neighbor approach and minimizes the entropy of the displaced frame differences.

2.4 Frequency domain methods

Transforming the BC assumption Eq.2.2 to Fourier domain yields

$$I^*(\omega_x, \omega_y, \omega_t) = I^*(\omega_x, \omega_y, \omega_t)e^{-j(v_x\partial t\omega_x + v_y\partial t\omega_y + \partial t\omega_t)} \tag{2.9}$$

where $I^*(\omega_x, \omega_y, \omega_t)$ is the Fourier transform of $I(x, y, t)$ and w_x, w_y, w_t denote spatio-temporal frequency. Clearly, for this equation to hold, it must satisfy

$$v_x\omega_x + v_y\omega_y + \omega_t = 0 \tag{2.10}$$

This is the basic constraint for frequency domain approaches. It states that all non-zero energy associated with a translating 2-D pattern lies on a plane through the origin in the frequency space, and the norm of the plane defines the optical flow vector. Frequency domain approaches are often presented as biological models of human motion sensing. They can handle cases that are difficult for matching approaches. But, extracting the non-zero energy plane usually involves heavy computation. As a consequence, they are not as popular as the other two types of approaches. Another strategy to improve robustness to brightness changes is to rely on phase correlation [25]. A recent version of this approach uses the Fourier-Mellin transform to calculate the phase correlation of each pair of co-sited patches [34].

2.5 Middlebury benchmark site: Top ranked methods

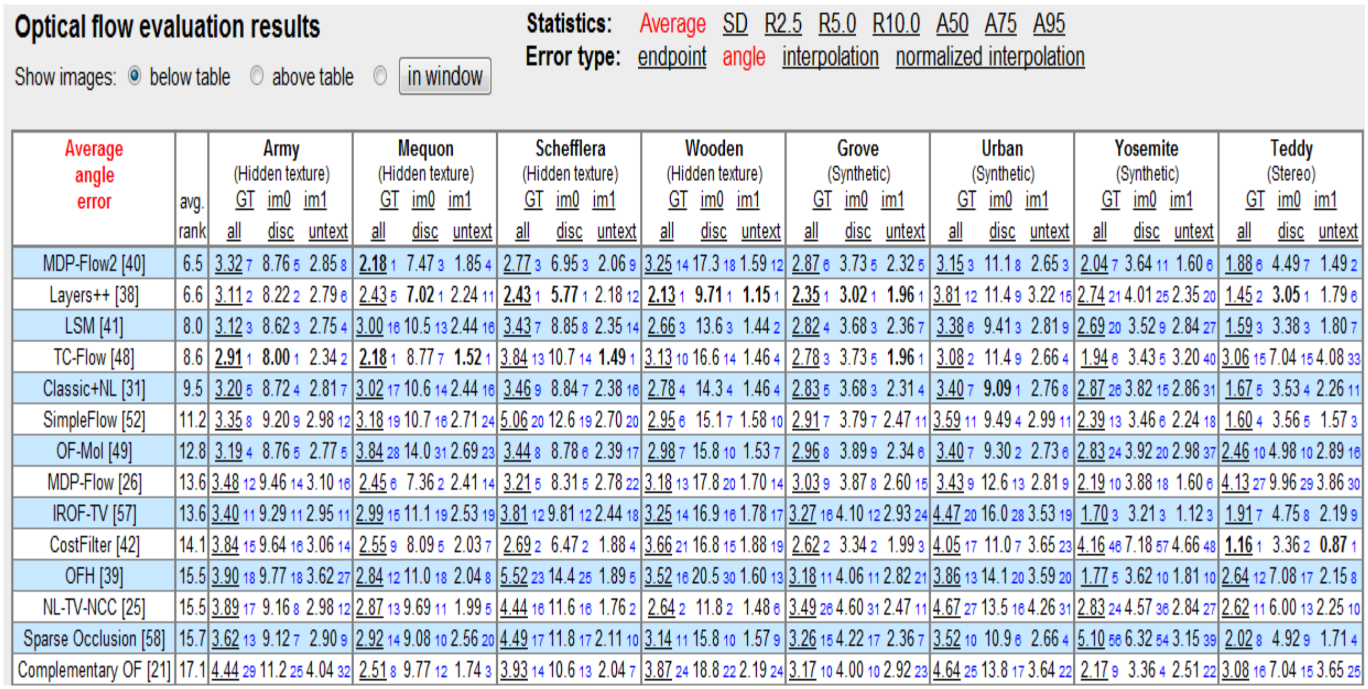
The Middlebury benchmark site [1] presents a ranked list of methods for OF computation. Of these, the top ranking methods are: The MDP Flow2 method [87] which estimates a coarse to fine motion flow for both large and small displacements; the Classic+NL method [79] which employs modern optimization techniques to improve the accuracy of the classical Horn-Schunk approach [35]; the Layers++ method [80] that employs a layered model-based approach aimed at overcoming shortcomings of previous approaches via a probabilistic OF estimator. The BC assumption is however common to all of the above mentioned methods. For handling large motion scenarios, a concept of coarse-to-fine image warping is introduced in [50]. The flow estimates are initialized from coarser levels, where displacements are small enough to be estimated by local optimization. This procedure cannot estimate the flow of structures that are smaller than their displacement. Brox et al [13] address this problem by integrating the energy minimization and descriptor matching strategies into a variational approach and guide the local optimization to large displacement solutions.

2.6 Large displacement OF methods

For handling large motion scenarios, a concept of coarse-to-fine image warping is introduced in [50]. They initialize the flow estimates from coarser levels, where displacements are small enough to be estimated by local optimization. This procedure cannot estimate the flow of structures that are smaller than their displacement. Brox et al [13] address this problem by integrating the energy minimization and descriptor matching strategies into a variational approach and guide the local optimization to large displacement solutions.

Matching by maximising mutual information is quite common in the medical image registration methods [66]. These methods estimate the sparse transformation field by aligning the centres of large size image patches. Hence, these methods cannot be used directly to compute dense flow fields on normal images. Glocker's work [27] for instance, uses a MRF model for non-rigid registration with the optimal MRF found by maximising a similarity measure such as mutual information and normalized mutual information etc, computed on relatively large size patches. Mutual information is known to be non-convex and has typically many local maxima [59, 29]. Therefore, the non-convexity and hence non-linearity of the motion estimation problem is enhanced by the usage of mutual information. Finally, mutual information decouples the brightness value from the location information. Hence, judging the output of the flow estimation process is difficult.

This thesis attempts to show that a *local information* constancy assumption acts as a more general criterion that overcomes many limitations of the methods discussed above and hence propose a local-level information maximization criteria for OF estimation.



Optical flow evaluation results

Show images: below table above table in window

Statistics: Average SD R2.5 R5.0 R10.0 A50 A75 A95
Error type: endpoint angle interpolation normalized interpolation

Average angle error	avg.	Army (Hidden texture) GT im0 im1			Mequon (Hidden texture) GT im0 im1			Schefflera (Hidden texture) GT im0 im1			Wooden (Hidden texture) GT im0 im1			Grove (Synthetic) GT im0 im1			Urban (Synthetic) GT im0 im1			Yosemite (Synthetic) GT im0 im1			Teddy (Stereo) GT im0 im1		
	rank	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext
MDP-Flow2 [40]	6.5	3.32 7	8.76 5	2.85 8	**2.18** 1	7.47 3	1.85 4	2.77 3	6.95 3	2.06 9	3.25 14	17.3 18	1.59 12	2.87 6	3.73 5	2.32 5	3.15 3	11.1 8	2.65 3	2.04 7	3.64 11	1.60 6	1.88 6	4.49 7	1.49 2
Layers++ [38]	6.6	3.11 2	8.22 2	2.79 6	2.43 5	**7.02** 1	2.24 11	**2.43** 1	**5.77** 1	2.18 12	**2.13** 1	**9.71** 1	**1.15** 1	**2.35** 1	**3.02** 1	**1.96** 1	3.81 12	11.4 9	3.22 15	2.74 21	4.01 25	2.35 20	1.45 2	**3.05** 1	1.79 6
LSM [41]	8.0	3.12 3	8.62 3	2.75 4	3.00 16	10.5 13	2.44 16	3.43 7	8.85 8	2.35 14	2.66 3	13.6 3	1.44 2	2.82 4	3.68 3	2.36 7	3.38 6	9.41 3	2.81 9	2.69 20	3.52 9	2.84 27	1.59 3	3.38 3	1.80 7
TC-Flow [48]	8.6	**2.91** 1	**8.00** 1	2.34 2	**2.18** 1	8.77 7	**1.52** 1	3.84 13	10.7 14	**1.49** 1	3.13 10	16.6 14	1.46 4	2.78 3	3.73 5	**1.96** 1	3.08 2	11.4 9	2.66 4	1.94 6	3.43 5	3.20 40	3.06 15	7.04 15	4.08 33
Classic+NL [31]	9.5	3.20 5	8.72 4	2.81 7	3.02 17	10.6 14	2.44 16	3.46 9	8.84 7	2.38 16	2.78 4	14.3 4	1.46 4	2.83 5	3.68 3	2.31 4	3.40 7	**9.09** 1	2.76 8	2.87 26	3.82 15	2.86 31	1.67 5	3.53 4	2.26 11
SimpleFlow [52]	11.2	3.35 8	9.20 9	2.98 12	3.18 19	10.7 16	2.71 24	5.06 20	12.6 19	2.70 20	2.95 6	15.1 7	1.58 10	2.91 7	3.79 7	2.47 11	3.59 11	9.49 4	2.99 11	2.39 13	3.46 6	2.24 18	1.60 4	3.56 5	1.57 3
OF-Mol [49]	12.8	3.19 4	8.76 5	2.77 5	3.84 28	14.0 31	2.69 23	3.44 8	8.78 6	2.39 17	2.98 7	15.8 10	1.53 7	2.96 8	3.89 9	2.34 6	3.40 7	9.30 2	2.73 6	2.83 24	3.92 20	2.98 37	2.46 10	4.98 10	2.89 16
MDP-Flow [26]	13.6	3.48 12	9.46 14	3.10 16	2.45 6	7.36 2	2.41 14	3.21 5	8.31 5	2.78 22	3.18 13	17.8 20	1.70 14	3.03 9	3.87 8	2.60 15	3.43 9	12.6 13	2.81 9	2.19 10	3.88 18	1.60 6	4.13 27	9.96 29	3.86 30
IROF-TV [57]	13.6	3.40 11	9.29 11	2.95 11	2.99 15	11.1 19	2.53 19	3.81 12	9.81 12	2.44 18	3.25 14	16.9 16	1.78 17	3.27 16	4.10 12	2.93 24	4.47 20	16.0 28	3.53 19	1.70 3	3.21 3	1.12 3	1.91 7	4.75 8	2.19 9
CostFilter [42]	14.1	3.84 15	9.64 16	3.06 14	2.55 9	8.09 5	2.03 7	2.69 2	6.47 2	1.88 4	3.66 21	16.8 15	1.88 19	2.62 2	3.34 2	1.99 3	4.05 17	11.0 7	3.65 23	4.16 46	7.18 57	4.66 48	**1.16** 1	3.36 2	**0.87** 1
OFH [39]	15.5	3.90 18	9.77 18	3.62 27	2.84 12	11.0 18	2.04 8	5.52 23	14.4 25	1.89 5	3.52 16	20.5 30	1.60 13	3.18 11	4.06 11	2.82 21	3.86 13	14.1 20	3.59 20	1.77 5	3.62 10	1.81 10	2.64 12	7.08 17	2.15 8
NL-TV-NCC [25]	15.5	3.89 17	9.16 8	2.98 12	2.87 13	9.69 11	1.99 5	4.44 16	11.6 16	1.76 2	2.64 2	11.8 2	1.48 6	3.49 26	4.60 31	2.47 11	4.67 27	13.5 16	4.26 31	2.83 24	4.57 36	2.84 27	2.62 11	6.00 13	2.25 10
Sparse Occlusion [58]	15.7	3.62 13	9.12 7	2.90 9	2.92 14	9.08 10	2.56 20	4.49 17	11.8 17	2.11 10	3.14 11	15.8 10	1.57 9	3.26 15	4.22 17	2.36 7	3.52 10	10.9 6	2.66 4	5.10 56	6.32 54	3.15 39	2.02 8	4.92 9	1.71 4
Complementary OF [21]	17.1	4.44 29	11.2 25	4.04 32	2.51 8	9.77 12	1.74 3	3.93 14	10.6 13	2.04 7	3.87 24	18.8 22	2.19 24	3.17 10	4.00 10	2.92 23	4.64 25	13.8 17	3.64 22	2.17 9	3.36 4	2.51 22	3.08 16	7.04 15	3.65 25

Figure 2.1 Screenshot of the middlebury optical flow evaluation and ranks web page

Chapter 3

INFORMATION FLOW

3.1 Motivation

Our motivation is to find a solution for *robust* estimation of OF. In a motion sequence, there is indeed an apparent variation in the brightness pattern across frames. In the past, this has been taken literally to be the source of the flow field and accordingly spatiotemporal intensity gradients have formed the basis for OF estimation. However, humans perceive motion robustly under varying ambient conditions (illumination, noise).

A remarkable example of this is the second order, non-Fourier type of motion perception demonstrated by Sperling [76] and texture patterns in [2]. The distinguishing feature of this type of motion is that the patterns consist of regions of microtextures which move coherently. Clearly, the BC assumption does not hold here and hence it cannot account for the perceived motion between regions. Texture-information, which is a higher order statistic than luminance information, is said to be exploited in the perception this type of motion [19].

Taking cue from this, we argue that underlying the apparent variation in brightness pattern is a flow of local information across frames. Accordingly, we formulate OF as an apparent movement of *local information* between the frames rather than of brightness patterns. This formulation removes the direct dependence of OF on flow of brightness patterns and hence, permits build in an inherent robustness to distortions such as illumination changes, noise, etc., which preserve the statistical structure of the images. By this formulation, if we consider a pixel in frame, local information given by this pixel flows to another unknown pixel of the next frame. Thus, the OF estimation can be modelled as a problem of estimation of motion of local information between two pixels across successive frames. We propose using a *divergence* metric to identify the corresponding pixel in the adjacent frame and compute the flow vector.

Divergence based approaches are adopted in registering multimodal and multiband images in medical and remote sensing applications [53, 66, 78] where the modalities can have very different contrast and hence brightness or pixel intensity is not a useful feature. The Kullback-Leibler's divergence is popular in estimating the transformation function that relates the images to be registered. The degree of change (transformation) between the two images to be registered is generally gross in comparison to that expected in successive frames of a video. Hence, it is of interest to examine if a divergence based strategy can be adopted for accurate OF computation. The key issues of interest here are the effect of this strategy on the accuracy/robustness of the extracted flow field and the ability to handle different magnitudes of displacements between frames.

3.2 Optical flow to Information flow

It is often advantageous to treat images as realizations of a stochastic process as it helps quantify questions regarding image information content via probability distributions and correlation functions [75]. Examples of such an approach can be seen in medical image registration, visual context modelling, etc.

3.2.1 Constant local information constraint

We assume a frame, denoted as $I(\mathbf{x})$ with **x** being the location vector, in a given image sequence to be a stochastic process and consider patches centered around each pixel in the frame as a set of events. Hence, a local patch around a pixel at **x** represents the random variable associated with the pixel and this local patch is constituted by pixels in the neighborhood $\Omega(\mathbf{x})$. We assume that pixels in a local patch to be samples of the random variable associated with the pixel at **x**. For convenience, $\Omega(\mathbf{x})$ is taken to be a block of size SxS centered at **x**. The random variable associated with the pixel at **x** is denoted as

$$R(\mathbf{x}) = \{I(\mathbf{y})|\mathbf{y} \in \Omega(\mathbf{x})\} \tag{3.1}$$

The conventional BC assumption is expressed as $I(\mathbf{x} + \Delta\mathbf{x}, t + \Delta t) = I(\mathbf{x}, t)$ with I denoting the brightness at **x** and t denoting time. This is used to establish corresponding pixels in successive frames. We introduce an analogous constant-local-information constraint for finding the corresponding pixels based on the match in the information present in them:

$$Inf(R(\mathbf{x} + \Delta\mathbf{x}, t + \Delta t)) = Inf(R(\mathbf{x}, t)) \tag{3.2}$$

where $Inf(R(\mathbf{x}))$ denotes information present in $R(\mathbf{x})$. This constraint states that the motion $\Delta\mathbf{x}$ should not change the information. Thus, the information measure for the pixel at **x** is invariant to motion between a frame pair. This is more flexible than the BC constraint as it tolerates some local deformation. If the measures are carefully selected and designed, it is also robust to any distortion like illumination fluctuations.

3.2.2 Information Flow Constraint Equations

Two variants of the proposed formulation in Eq.3.2 arise from considering two kinds of information measures. For the first variant, we can consider the self information of one RV ($R(\mathbf{x})$) while for the second one, we can consider the information in two RVs (in two frames).

3.2.2.1 Self-information measures

In information theory, self-information is a measure of the information content associated with the outcome of a random variable. The term self-information is also sometimes used as a synonym of entropy and was introduced by Claude E. Shannon [74]. An example of entropy optimization based motion estimation is the approach in [11]. A systematic attempt to develop a generalization of Shannon's entropy was carried by out Renyi [69], who characterized an entropy of order k as

$$H_k(R) = (1-k)^{-1} \log(\sum_{i=1}^{n} P(r_i)^k), \;\; k \neq 1, \; k > 0 \tag{3.3}$$

where R is a random variable, P its probability density function (pdf) and k is a real number. We can easily verify that

$$\lim_{k \to 1} H_k(R) = H(R) \tag{3.4}$$

where $H(R)$ is the Shannon's entropy.

Coming back to Eq.3.2, when we use entropy as an information measure,

$$H_k(R(\mathbf{x} + \Delta\mathbf{x}, t + \Delta t)) - H_k(R(\mathbf{x}, t)) = 0 \tag{3.5}$$

This equation is the equivalent of the well known BC equation. It can be interpreted as imposing BC-like constraint on the patch around $\mathbf{x}$ in the entropy space before and after motion. Using a first order approximation for the first term, it is clear that

$$\nabla_{\mathbf{x}}^T H_k(R(\mathbf{x}, t))\Delta\mathbf{x} + \nabla_t H_k(R(\mathbf{x}, t))\Delta t = 0$$
$$\nabla_{\mathbf{x}}^T H_k(R(\mathbf{x}, t))\tfrac{\Delta\mathbf{x}}{\Delta t} + \nabla_t H_k(R(\mathbf{x}, t)) = 0$$

$$\nabla_{\mathbf{x}}^T H_k(R(\mathbf{x}, t))\mathbf{v} + \nabla_t H_k(R(\mathbf{x}, t)) = 0 \tag{3.6}$$

where $\mathbf{v} = [\frac{\Delta\mathbf{x}}{\Delta t}]$ is the flow vector at $\mathbf{x}$; $\nabla_{\mathbf{x}}^T H_k(R(\mathbf{x}, t))$ is the gradient, and $\nabla_t H_k(R(\mathbf{x}, t))$ is the entropy frame difference. We call the Eq.3.6 as **Local information flow constraint**.

Denoting the term $\nabla_{\mathbf{x}}^T H_k(R(\mathbf{x}, t))$ as $\mathbf{U}$ and the term $\nabla_t H_k(R(\mathbf{x}, t))$ as $\mathbf{b}$, we see that Eq.3.6 is a simple linear constraint on motion.

$$\mathbf{U}\mathbf{v} = -\mathbf{b} \tag{3.7}$$

If the $rank(\mathbf{U}) = 2$, the least squares solution solution, $\mathbf{v} = -(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{b}$, of this overdetermined system leads to a unique determination of the information flow at $\mathbf{x}$, and thus the motion at $\mathbf{x}$.

3.2.2.2 Mutual information based measures

If we consider a pair of RVs (each from a pair of successive frames) the amount of information one random variable (in frame1) contains about the other (in frame2) is of relevance. In Eq.3.2, there are two RVs; one on left hand side ($R(\mathbf{x} + \Delta\mathbf{x}, t + \Delta t)$) and a second one on the right hand side ($R(\mathbf{x}, t)$). Reformulating this with mutual information (MI) as an information measure the goal of OF estimation is maximisation of this information.

$$maximize\ MI(R(\mathbf{x} + \Delta\mathbf{x}, t + \Delta t); R(\mathbf{x}, t)) \tag{3.8}$$

$$\Delta\mathbf{x} = argmax_{\Delta\mathbf{x} \in S_r}(MI(R(\mathbf{x} + \Delta\mathbf{x}); R(\mathbf{x})))$$

where MI is a measure of mutual information and $\Delta\mathbf{x}$ is the unknown flow vector describing the movement of a pixel at **x** between successive frames and S_r is the search range.

One way to define the mutual information associated with a pair of RVs is via the *divergence* measure [44]. The divergence between a pair of of RVs is defined as the distance between the product of their marginal pdfs and the joint pdf. Such measures have been used in the analysis of contingency tables [28], in signal processing [40, 41], in pattern recognition [7, 18] and in medical image registration [54, 31, 55, 22]. These are based on generalised divergence theories and are known to offer improved accuracy, robustness and speedy convergence for image registration [54, 22]. We next, present some of the divergence measures that are of interest.

Kullback-Leibler Divergence (KLD):
The canonical measure for mutual information was introduced by Shannon and Weaver in 1949. It is also known as Kullback-Leibler Divergence [44]. Given two random variables R_1 and R_2 the KLD between them is defined as the distance between the joint pdf and the product of the marginal pdfs. It is expressed as

$$KLD(R_1; R_2) = \sum_{r_1 \in R_1} \sum_{r_2 \in R_2} P_{R_1R_2}(r_1, r_2) log(\frac{P_{R_1R_2}(r_1, r_2)}{P_{R_1}(r_1)P_{R_2}(r_2)}). \tag{3.9}$$

where R_1 and R_2 are two patches in two video frames; $P_{R_1R_2}(r_1, r_2), P_{R_1}(r_1), P_{R_2}(r_2)$ are the joint and marginal probability density functions (pdf) of random variables R_1 and R_2 respectively; and r_1, r_2 are realizations of random variables R_1 and R_2 respectively.

Normalized Kullback-Liebler Divergence (NKD):
In medical image registration, KLD is generally normalized with the corresponding average entropy [23]. This is called as Normalized Kullback-Liebler Divergence measure (D) was used for image registration in [64]. NKD is defined in [56] as

$$NKD(R_1; R_2) = \frac{-2 \sum_{r_1 \in R_1} \sum_{r_2 \in R_2} P_{R_1R_2}(r_1, r_2) log(\frac{P_{R_1R_2}(r_1, r_2)}{P_{R_1}(r_1)P_{R_2}(r_2)})}{\sum_{r_1 \in R_1} P_{R_1}(r_1) log(P_{R_1}(r_1)) + \sum_{r_2 \in R_2} P_{R_2}(r_2) log(P_{R_2}(r_2))}. \tag{3.10}$$

Kolmogorov Divergence (KD):
Recently, an attempt has been made to increase the computational tractability of divergence measures based on Kolmogorov's divergence theory [90]. This results in a measure which is defined as the Manhattan distance between two distributions of RVs R_1 and R_2.

$$KD(R_1; R_2) = \sum_{r_1 \in R_1} \sum_{r_2 \in R_2} |P_{R_1 R_2}(r_1, r_2) - P_{R_1}(r_1) P_{R_2}(r_2)|. \tag{3.11}$$

It can be noted that this essentially measures the degree of independence between R_1 and R_2.

I_α-Divergence (ID):
Estimation of I_α-information arises as a step towards non-parametric estimation of Shannon entropy. However, estimation of the α-entropy is of interest in its own right and has been employed for image retrieval [32]. This has been used for image registration from multiple modalities via the α-Jensen difference [52]. The expression for ID is given as

$$ID(R_1; R_2) = \frac{1}{\alpha(\alpha - 1)} (\sum_{r_1 \in R_1} \sum_{r_2 \in R_2} \frac{P^{\alpha}_{R_1 R_2}(r_1, r_2)}{P^{\alpha-1}_{R_1}(r_1) P^{\alpha-1}_{R_2}(r_2)} - 1). \tag{3.12}$$

Renyi Divergence (RD):
Another tractable divergence measure is the RD measure [33] as it requires only marginal pdf estimation and avoids both polynomial expansions and estimation of joint pdf. The RD measure for a pair of RVs is given by

$$RD(R_1; R_2) = \frac{1}{\alpha - 1} \log \sum_{r_1 \in R_1} \sum_{r_2 \in R_2} P^{\alpha}_{R_1}(r_1) P^{1-\alpha}_{R_2}(r_2). \tag{3.13}$$

Tsallis Divergence (TD):
Tsallis divergence [81] is a generalization of the standard Boltzmann-Gibbs entropy. A variant of this divergence measure is reported as TD measure which is as given as

$$TD(R_1; R_2) = \frac{1}{(\alpha - 1)} (\sum_{r_1 \in R_1} \sum_{r_2 \in R_2} \frac{P^{\alpha}_{R_1 R_2}(r_1, r_2)}{P^{\alpha-1}_{R_1}(r_1) P^{\alpha-1}_{R_2}(r_2)} - 1). \tag{3.14}$$

Our proposal is to use Eq.3.8 to estimate motion by matching pixels across frames. The information measure to be used can be any of the variants of the divergence measures listed above, which are generally viewed as inversely proportional to the similarity between 2 RVs. Since the variants presented above estimate the divergence between joint and marginal pdfs, maximization of $MI(R_1; R_2)$ serves to maximize the dependence of one RV on the other.

The divergence variants are derived from statistical information and not brightness. This aspect distinguishes them from other similarity measures: i) a divergence measure has the capacity to capture any kind of relationship between variables because it is built from the joint and marginal pdfs of the variables and does not utilize statistics of any grade or order; ii) it is shift invariant under spatial transformations [44]. The second property is based on the fact that the argument of the summands in the above equations are non-dimensional [26]. These properties hold for all kind of motions, such as translations, rotations, and any movement that preserves the order of the original elements of the variables. Thus, the proposed estimator is a robust OF estimator.

3.3 Implementation issues

A number of issues arise in the course of implementation of the proposed approach for OF estimation. We discuss them next.

3.3.1 Block size and Boundary issues

In the proposed formulation, a patch around *every* pixel is considered for computing local information. The patch is chosen as a block of size SxS. S is an important parameter in this algorithm just as it is in phase correlation or block matching based motion estimation methods. S should be large enough to capture large displacement vectors and small enough so that these vectors remain constant within the block. Empirically, it was found that a block size of 11x11 provided the best results for an image size of around 300x500.

In practice, a fixed block size of 11x11 cannot be maintained at the image boundaries. Hence, the block size was varied depending on the distance from the image boundaries. Suppose l,r,t and b are the distances of a pixel from the left, right, top and bottom boundaries respectively. Then the block size was chosen to be mxn where $m = min(l, 5) + min(r, 5) + 1$ and $n = min(t, 5) + min(b, 5) + 1$.

3.3.2 Range of displacement

The choice of the range of displacement (or the search range for match) impacts a few factors, with the key ones being computational efficiency and ability to handle small versus large displacements. In video sequences acquired at a good frame rate, the displacement of a pixel between two adjacent frames is no more than 5 pixels. Hence, the search range for finding the matching patch was fixed to [-7:7]. This can be made higher for handling larger displacements however it will incur a higher computational load.

3.3.3 Estimation of Probability Density Functions

Since the proposed approach is dependent on an information measure, PDF estimation is required. This is a fundamental yet difficult problem in computer vision. Most published methods can be classified as either histogram based methods or kernel density estimation (KDE, also called Parzen Windows estimators) based methods. The latter has the advantage of being nonparametric. However, both have limitations. As a consequence of the law of large numbers, histograms require a relatively large number of observed pixels to give an accurate and smooth estimate of the underlying density. Hence, they perform poorly for the local estimation (from small image patches) of information measures. KDEs require relatively fewer samples to a give a good density estimate, but the result critically depends upon choosing the window width of the kernel. An excellent overview of nonparametric PDF estimators can be found in [39].

In our work, a multi-variate Parzen kernel density estimation [85, 84] was used to estimate the pdfs. This involves placing a kernel function at each sample location and evaluating the density as a sum of the kernels. The estimation was based on a Gaussian kernel of bandwidth given as

$$\sigma_N = 2.345.\sigma_N^*.N^{-1/5} \tag{3.15}$$

where σ_N^* is the estimated standard deviation for N (=S^2) data points [9]. In this work, $S = 11$ was used.

3.3.4 Smoothing

In order to suppress noise due to the gradient terms in divergence based computation, a smoothing step was included: The velocity at **x** was computed as the weighted average of the velocities of its neighbors. The weights were chosen to be dependent on the information associated with the pixels. This non-linear weight mapping helps in forcing the motion vector to be in the direction of maximum information flow.

Let $\mathbf{v_x}$ be the estimated velocity for a pixel at **x**. Then, the smoothed velocity denoted as $\mathbf{v'_x}$ is computed as

$$\mathbf{v'_x} = \frac{\sum_{\mathbf{y}\in N_\mathbf{x}} MI(R(\mathbf{y}+\mathbf{v_y}); R(\mathbf{y}))\mathbf{v_y}}{\sum_{\mathbf{y}\in N_\mathbf{x}} MI(R(\mathbf{y}+\mathbf{v_y}); R(\mathbf{y}))} \tag{3.16}$$

where $N_\mathbf{x}$ is the 5x5 neighborhood around **x** and $MI(R(\mathbf{y}+\mathbf{v_y}); R(\mathbf{y}))$ is the mutual information between RVs associated with **y** and $\mathbf{y}+\mathbf{v_y}$. Sub-pixel accurate flow vectors are found following [77] by simply upsampling the input image using bicubic interpolation.

3.4 Summary

In this chapter, we have presented the concept of information flow and derivation of *information flow* constraint equations and also explained about two kinds of information measures. We also discussed about some issues in the implementation of proposed formulation.

Chapter 4

EXPERIMENTS & RESULTS

The performance of the proposed robust OF estimation algorithms was evaluated using the benchmark Middlebury database [1], which includes synthetic and real scenes. The database consists of two sets, one with ground truth for calibration of the method and a second one without ground truth. Both sets were used in the assessment. These data sequences provide challenges including the aperture problem, texture-less regions, motion discontinuities, occlusions, large motions, small objects, non-rigid motion, mixed pixels, changes in illumination, motion blur, non-Lambertian reflectance, and camera noise [5]. Thus, these sequences provide meaningful comparisons across OF algorithms.

Since the proposed method adopts a matching strategy based on divergence, to check if the accuracy of the flow computation is compromised we considered a weighted variant of the divergence term. For this purpose we followed [65] which uses a gradient vector based weighting approach. The weighting function is defined as

$$W(R_1, R_2) = \sum_{(\mathbf{x},\mathbf{x'})\in(R_1\cap R_2)} \cos^2(\alpha_{\mathbf{x},\mathbf{x'}}) \tag{4.1}$$

where $\alpha_{\mathbf{x},\mathbf{x'}}$ is the angle between the gradient vectors $\nabla\mathbf{x}$ and $\nabla\mathbf{x'}$ at a sample point (in one frame) and its corresponding point (in the next frame) and is defined as $\alpha_{\mathbf{x},\mathbf{x'}} = \arccos \frac{\nabla\mathbf{x}.\nabla\mathbf{x'}}{|\nabla\mathbf{x}||\nabla\mathbf{x'}|}$. The data term to be maxmized is defined as

$$D(R_1, R_2) = W(R_1, R_2)MI(R_1, R_2) \tag{4.2}$$

We first present a comparison of OF estimation obtained with Eq.3.8 and Eq.4.2. These are reported for different divergence measures. Table 4.1, lists the errors in estimated OF for the Rubberwhale image sequence [1]. As in [1], the vector difference **e** between the ground truth OF and estimated OF is determined and the average and standard deviation of the angle of this **e** as AAE and SAE, respectively, are reported. From the tabulated results it appears that, on average, the weighting helps reduce the errors by roughly 15 to 18$\%$. Hence, in the rest of our experiments we used weighted variants to evaluate the proposed formulation.

Computed flow maps for the proposed formulation along with the ground truth flow are shown for the Rubberwhale sequence in the Fig.4.1. All flow maps are color coded according to [5] (see Fig.4.2)

Table 4.1 Comparison of proposed OF estimation with and without weighting of the information measures for the Rubberwhale image sequence

MI measure	AAE	SAE
Without weighting		
KLD	4.48	20.6
NKD	3.63	9.20
KD	12.8	14.7
ID	9.65	16.9
RD	27.5	26.5
TD	13.8	24.7
Average error	**11.9**	**18.8**
With weighting		
KLD	4.98	9.97
NKD	3.19	7.41
KD	10.3	11.9
ID	10.1	17.4
RD	20.6	22.6
TD	11.4	23.2
Average error	**10.1**	**15.4**

plus scaled and resized. Included here is the flow map estimated with self information measure H, (mentioned in Section 3.2.2.1). It can be seen clearly from the figures that the results of NKD, H, and KLD coincide well with the ground truth flow field. (Please use soft copy of the thesis to observe colors of the figures.)

Next, we present results of a quantitative assessment of the computed flow fields computed for six sequences. Comparison of performance with some leading OF estimation methods are also included. Experiments were specifically done to test the accuracy, robustness to distortions of the flow field and computational speed of computation.

4.1 Accuracy

The proposed method uses local statistics to match at the pixel level to produce dense flow field. Hence, the accuracy of the proposed method was compared with the results of other existing dense flow algorithms such as Classic+NL[80], Black-Anandan method [10], TV-L1-Flow [86] and a modern Horn-Schunck's method [79]. The AAE and SAE figures are presented in Table 4.2 are taken from Middlebury website [1]. The best figures (least error) for each sequence are indicated in bold and the average rank (across the 12 columns) of a method is listed as part of the suffix in the label for the method in the first column.

From these figures, it can be observed that overall: a) the proposed formulation with the NKD measure performs competitively against other existing methods; b) most of the errors with the proposed method are at corner points (also seen in Fig.4.1) which could possibly be due to the final smoothing step in our algorithm; c) almost all the divergence variants outperform existing methods on the challenging

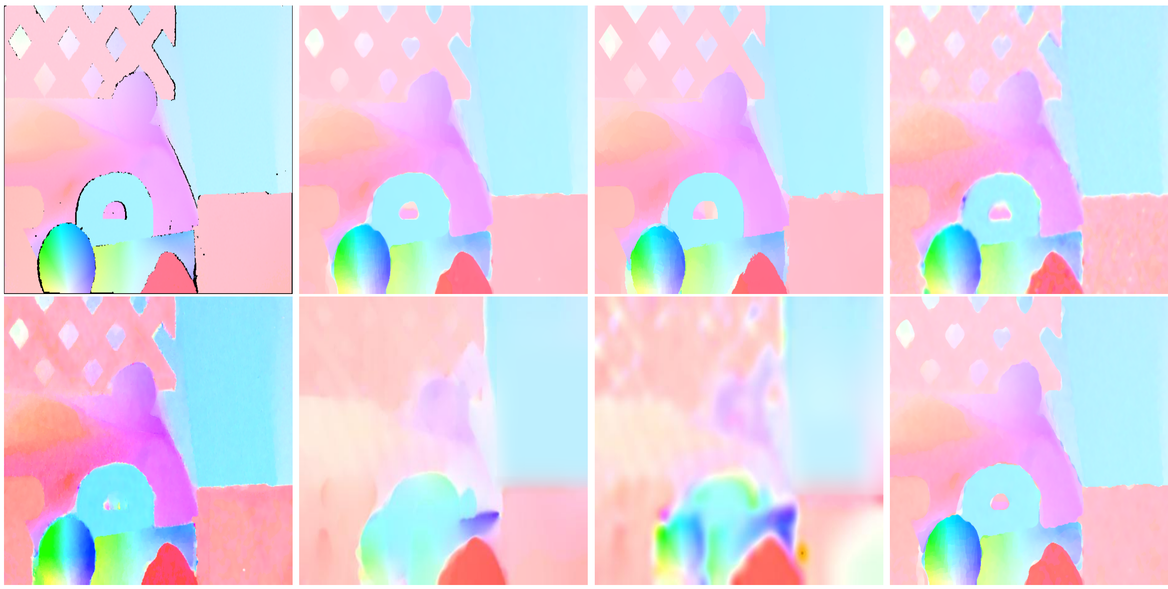


Figure 4.1 Results on Rubberwhale data sequence. Top row, left to right: Ground truth flow; Flow fields with KLD, NKD and KD. Bottom row, left to right shows flow fields with ID, RD, TD, H.

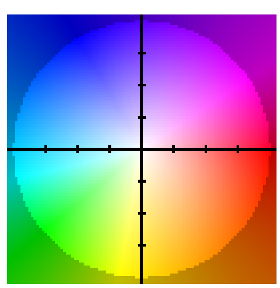


Figure 4.2 Color coding of flow vectors: Direction is coded by hue and magnitude by saturation.

Table 4.2 Comparison of AAE and SAE error metrics for different methods on the Middlebury data sequences.

Method$_{avg.rank}$	Army		Schefflera		Wooden		Grove		Urban		Yosemite	
	AAE	SAE	AAE	SAE	AAE	SAE	AAE	SAE	AAE	SAE	AAE	SAE
Classic+NL$_{1.83}$	3.20	**7.40**	3.46	12.3	2.78	**12.5**	**2.83**	5.03	3.40	8.94	2.87	**2.29**
TV-L1-Flow$_{4.42}$	3.36	7.64	6.50	21.3	3.80	17.5	3.34	5.25	5.97	14.5	3.57	3.38
Black-Anandan$_{6.75}$	6.81	10.5	13.0	26.1	8.29	20.8	4.18	6.29	6.19	17.1	3.63	4.06
Horn-Schunck$_{7.91}$	8.01	11.8	14.2	25.5	12.4	26.1	4.64	6.12	8.21	19.9	4.01	3.95
NKD$_{1.81}$	**3.19**	7.41	**3.44**	**12.1**	2.98	13.3	2.96	5.05	3.40	9.18	**2.83**	2.57
KLD$_{4.41}$	4.98	9.37	5.53	20.9	3.02	14.3	4.48	6.52	**3.27**	11.5	3.62	5.64
KD$_{6.25}$	10.9	11.0	9.01	24.9	10.6	15.9	4.81	**4.74**	12.1	**8.71**	4.54	3.29
ID$_{8.41}$	10.7	17.5	11.5	23.3	15.4	17.0	4.65	10.7	11.7	14.6	25.0	13.9
TD$_{9.41}$	18.0	13.2	23.6	12.3	26.5	16.2	27.8	19.9	15.9	23.4	21.7	24.7
RD$_{9.25}$	20.6	22.6	12.1	35.4	11.0	25.8	31.5	17.5	14.1	17.0	5.18	13.5
H$_{5.75}$	5.64	11.5	6.32	21.7	**2.61**	19.2	3.74	7.69	5.36	14.7	3.18	5.19

Urban and Schefflera image sequences. In particular, our method with NKD outperforms most of other methods on scenes with hidden texture such as Army and Schefflera and synthetic scenes such as Yosemite. However, on the Wooden sequence, which is also a hidden texture sequence, the performance dips marginally. Performance with KLD is also comparable with that of other methods for Army, Urban, Schefflera, Yosemite sequences. OF computed with Kolmogorov divergence (KD) has very less SAE for Grove sequence while with H the error is low for Wooden image sequence.

In general, the accuracy levels of our method with RD and KD are very low which is to be expected as these divergence measures are designed to decrease the computational load by compromising the accuracy.

The reason for superior performance of our method with NKD (1st rank among all variants of information measure) is the simultaneous maximization of information along with minimizing the entropy factors. This is due to the normalization by the uncertainty terms in the denominator in Eq.3.10. The success (2^{nd} rank) of the KLD based method is due to the non-dimensional nature of the argument of the summand. The self information measure H based method is 3rd ranked due, mostly to its capability to minimize the distance between two marginal distributions.

The accuracy levels of the KD based method is not consistent on all images as NKD and KLD. This is possibly due to the fact that the measure is simplistic. An interesting outcome is the superior performance of the self information H- based method over RD,ID and TD-based estimation. This could be due to fact that unlike H, the methods based on RD,ID and TD use ratio of exponents of pdfs to match the pixels and computation of these exponents and ratios yield noisy OF estimation. Some surprising results can also be noted from Table 4.2: a) The KD criterion is showing very less SAE compared to other methods on Wooden and Urban sequences which are synthetic ones. b) The same behaviour is shown by H criterion in AAE terms on Wooden image sequence which contains hidden texture. c) The KLD criterion is showing less AAE value on Urban sequence which is a synthetic image sequence. These behaviours are possibly due to the gradient weighting in Eq.4.2. At the time of submission, the proposed formulation labeled as OF-MOI, with weighted NKD measure stands 15^{th} and 19^{th} in the AAE and SAE rankings respectively [1].

Additional results are shown in Fig.4.3 to demonstrate the quality of computed flow fields computed with NKD and classic+NL methods. The flow fields have been scaled and resized for closer examination. It can be seen clearly from the figures that the flow maps estimated with the proposed formulation coincide well with the ground truth flow field.

Detailed comparison can be found on Middlebury website [1] for the weighted NKD-based flow estimation. Based on these results, it can be concluded that the information maximisation strategy does not compromise accuracy of OF estimation. The inclusion of a gradient-based weight for the information measure serves to marginally boost the performance.

4.2 Robustness to distortions

A key benefit of information flow-based OF estimation should be increased robustness to image-level distortions. This was tested by imposing different distortions on 2 input data sequences of which one was a real and other synthetic sequence. Five types of degradation were considered: salt pepper noise (10% of the frame), overexposure from a camera flash simulated by a uniform mean shift (boost all pixel values by 20), missing data simulated by introducing random black patches, combination of the previous 3 degradations and spatially varying Gaussian blur.

Sample degraded images for a RubberWhale (real) frame and their corresponding flow fields estimated with NKD are shown in Fig. 4.4. The flow fields were estimated using the Classic+NL[80], kNN-ME method [11] and FMT method [34]. The last two methods are block-based and designed to handle degradations. These results were compared with the flows estimated with our formulation. The computed error metrics for the RubberWhale and Grove3 (synthetic) sequences are presented in Tables 4.3 and 4.4 respectively.

Table 4.3 Assessment results for flow estimation on degraded RubberWhale (real) image sequence.

	Original		Noise		Overexposure		Missing patches		Combo		Non-uniform blur	
Method	AAE	SAE	AAE	SAE	AAE	SAE	AAE	SAE	AAE	SAE	AAE	SAE
kNNME	6.76	10.8	15.1	22.9	9.10	12.5	13.6	21.3	17.2	21.5	11.5	8.52
FMT	10.1	16.2	19.4	26.0	12.8	25.7	20.2	41.7	28.1	38.4	18.3	35.2
Classic+NL	3.81	7.49	12.1	21.7	6.82	6.72	12.6	19.1	19.1	23.56	5.44	15.7
KLD	3.48	9.69	4.71	10.5	5.84	10.5	5.59	9.56	6.00	11.4	4.99	10.7
NKD	2.73	7.52	2.87	8.84	3.85	10.6	3.51	8.15	4.46	9.21	4.16	9.43
KD	10.8	17.7	12.3	23.4	11.6	21.5	11.8	19.8	19.1	29.4	11.4	22.5
ID	10.5	16.9	10.9	18.8	11.3	17.8	12.2	19.6	15.5	23.8	10.8	19.7
RD	17.5	16.6	19.3	18.5	17.7	21.0	20.5	28.4	22.1	31.7	18.2	28.8
TD	23.8	14.5	25.8	26.0	26.3	27.1	24.9	22.4	33.0	23.9	26.4	29.7
H	16.7	11.4	18.1	12.1	18.3	14.5	13.6	16.2	26.0	19.1	18.9	16.5

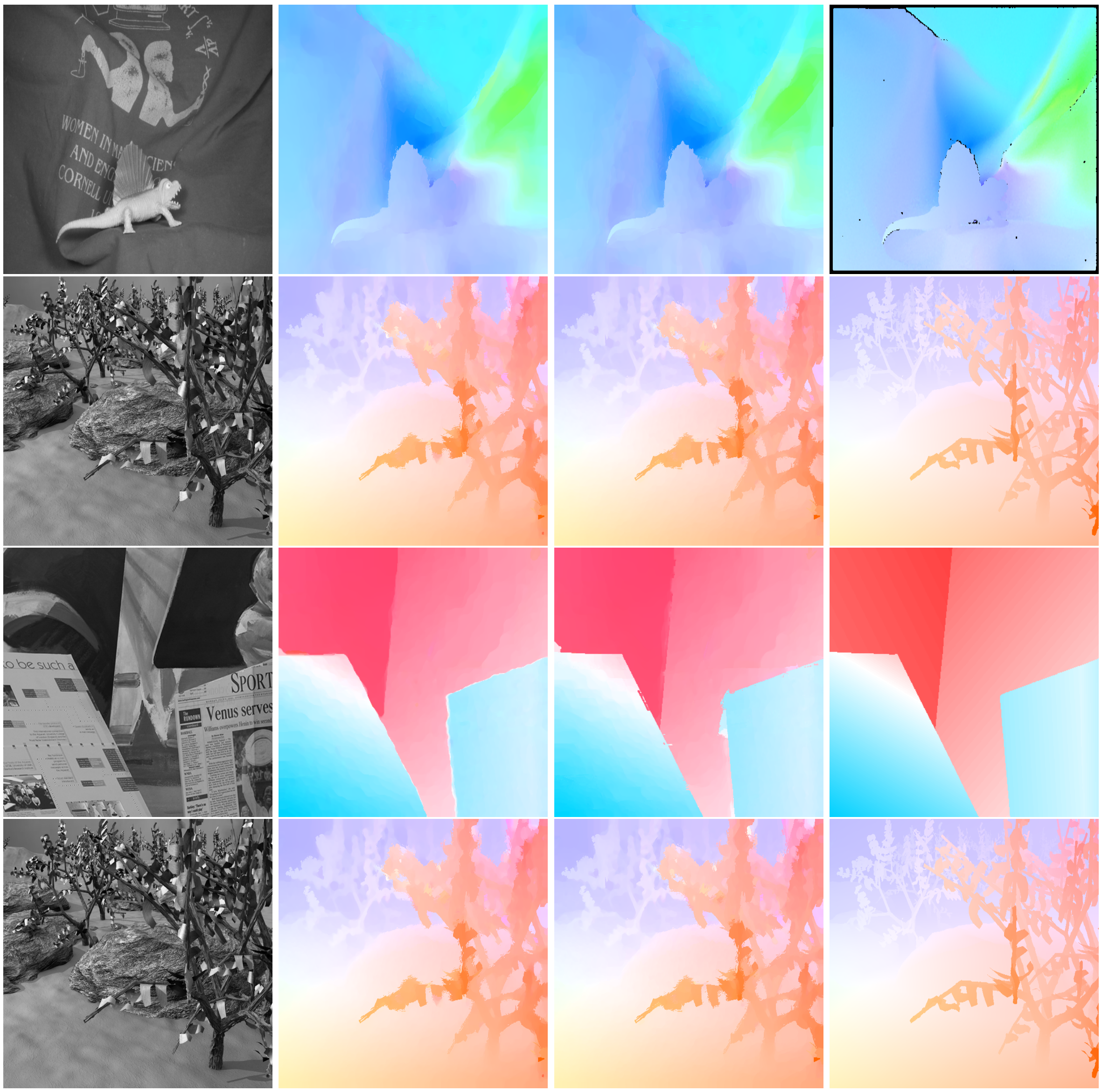


Figure 4.3 NKD results on some middlebury data sequences [Dimetrodon(hidden texture), Grove3 (synthetic) and Venus (stereo)] sequences: First column : frame 10, Second column: Our result, Third column: Result of classic+NL method, Fourth column: Groundtruth flow.

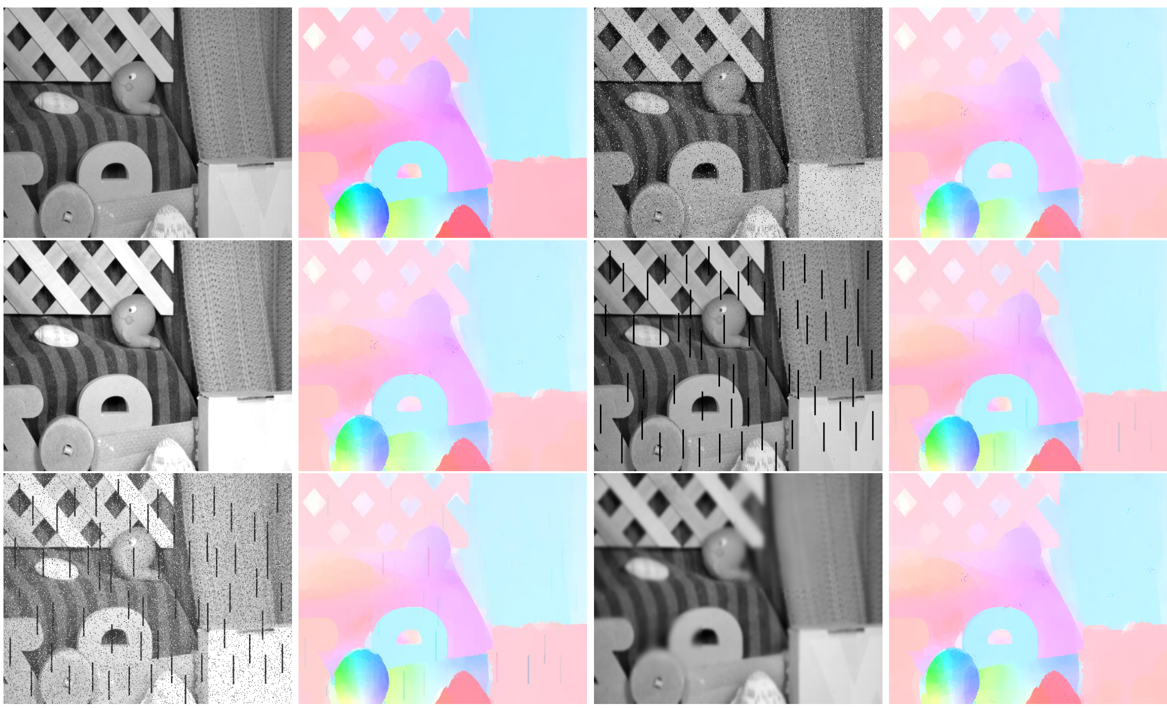


Figure 4.4 Flow fields for images with distortions. Top row: original image and a noisy frame; Second row: Overexposure and missing patches; Third row: multiple distortions and non-uniform blur

Table 4.4 Assessment results for flow estimation on degraded on Grove3 (synthetic) image sequence.

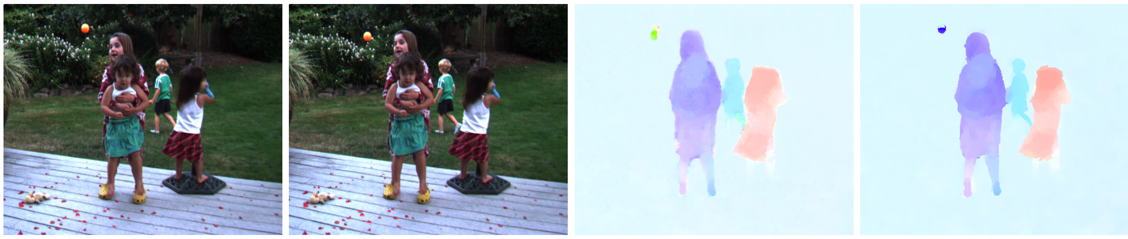
Method	Original		Noise		Overexposure		Missing patches		Combo		Non-uniform blur	
	AAE	SAE	AAE	SAE	AAE	SAE	AAE	SAE	AAE	SAE	AAE	SAE
kNNME	5.84	10.5	12.4	26.7	11.9	16.9	13.0	28.2	20.4	30.2	15.5	18.4
FMT	11.6	9.58	16.8	15.1	13.9	11.5	17.7	19.2	26.9	29.5	18.4	22.1
Classic+NL	3.38	5.62	15.6	16.8	7.11	9.75	22.1	9.16	22.7	26.8	10.9	9.57
KLD	4.79	3.17	6.50	6.51	5.18	9.51	5.75	8.73	7.85	12.3	5.25	7.08
NKD	2.70	5.76	3.15	7.16	3.94	5.96	5.16	9.41	7.29	11.5	4.82	6.73
KD	5.27	7.98	8.50	7.78	7.24	12.8	9.44	6.38	19.7	8.79	6.61	8.45
ID	5.74	18.3	7.38	15.5	6.81	19.7	9.65	9.95	10.9	12.7	5.86	8.17
RD	22.9	11.3	22.4	25.7	20.4	12.3	21.0	12.1	31.5	20.3	26.7	12.6
TD	25.8	17.3	25.6	19.1	27.1	14.5	33.5	20.4	40.5	21.0	27.4	19.8
H	5.58	16.1	6.85	8.13	7.23	14.8	7.51	10.5	10.0	23.2	11.9	14.9

The results in Table 4.3 indicate that the block-based approaches fare better under distortion compared to the Classic+NL method as the error increases rapidly for the latter, which is to be expected as these methods have an inherent resistance to noise as against pixel-level methods. However, the kNN-ME method [11] is susceptible to mixed, noisy, and blurry environments while the FMT method [34] is relatively robust to only brightness changes and blur. The proposed formulation with all variants perform consistently well across sequences as the increase in error under distortions is quite small. This confirms that the information maximisation criterion is indeed a good strategy for robust OF estimation. The results on the synthetic sequence exhibit a trend similar to that on real sequence.

4.3 Large Motion

Classical OF estimation requires dense sampling in time. But, in real scenarios, there may be videos (surveillance for instance) with low sampling rate. Classical approaches result in erroneous OF estimation in such scenarios. However, coarse-to-fine warping schemes have relaxed this constraint to some extent. Articulated motion in general and human motion in particular remain challenging. For instance, small body parts like hands can move extremely fast. OF estimation becomes more ambiguous in those regions. We have tested our proposed approach on the Backyard (Dog dance) image sequence [1] which have large displacements. The proposed formulation was used to estimate the OF by expanding the search range to [-25:25].

Brox et.al [12] proposed a method tailored to handle large displacements. For comparison we show the flow fields estimated by that method and our (NKD-based) method in Fig.4.5. It is difficult to judge the accuracy levels of these flow fields as ground truth is not available. The flow fields of the proposed method (with NKD) and Brox's method appear to be similar. However, the structure of the boy in the background and the ball in the air appear to crisper in the flow field estimated by the proposed method. However, the computation times for the two methods are not comparable: 94 seconds for [12] and 1763 seconds for our method.



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

Figure 4.5 OF estimation on frames with large displacement: a)Frame 07; b)Frame 08; c) Flow field with [12] and d) Flow field of the proposed method.

From the above qualitative and quantitative results of different experiments, it can be clearly seen that the proposed method is accurate, able to handle small and large displacements and robust under distortions.

4.4 Computational speed

Generally, local motion integration at a large and adaptive scale incurs a high computational cost. Computation of any information measure is expensive as it relies on joint and/or marginal pdf estimation. The computation time for our formulation with different variants along with other existing methods are presented in Table 4.5 on the Urban image sequence [5]. The reported computation times are for implementations on a dual core 1.83 GHz processor with 3GB of memory, on Windows 7 operating system, without any optimization of codes. We have used the available codes for some existing methods. Computing the statistics of a block of size 11x11 around every pixel and matching by searching within range of [-7:7] leads to a high computational load in our formulation. A fast searching algorithm such as diamond search algorithm or simulated annealing techniques etc, could be explored to reduce this cost. The local motion integration at a large and adaptive scale is another reason for the load. These deserve attention in the future. Interestingly, a leading, recent method [80] which uses a probabilistic OF model in layers is also computationally quite intensive.

Based on our extensive experiments the following recommendations can be made for flow computation with the proposed formulation: NKD for good *accuracy*; NKD, KLD for *robust* flow computation and H, RD for fast computation. Thus the local information-based approach offers both robustness and accuracy albeit at a higher computational load. A fast computation of flow is feasible with H with a small trade off in accuracy. While high computational cost may seem prohibitive for real time applications, the strength of our formulation in terms of robustness to degradations and ability to extract

Table 4.5 Comparison of computational times on Urban image sequence

Method	Run time(in secs)
TV-L1-Flow [86]	8
Horn-Schunck [35]	49
FMT [34]	814
kNN-ME [11]	905
Classic+NL [79]	972
Layers++ [80]	18206
RD	93
H	135
KD	516
KLD	1072
ID	1080
NKD	1120
TD	1374

structure from motion even from noisy data is of interest, especially in applications such as echocardiography. We will demonstrate some of these applications in the next section.

4.5 Summary

We have taken an information theoretic approach to OF computation and relied on statistical information to obtain an accurate and robust flow. The performance of the information-based formulation with NKD and KLD was found to be on par with the leading methods in terms of accuracy with an additional key strength of providing robustness to flow field estimation which current methods lack. Thus the information theoretic approach offers both robustness and accuracy without comprising either of them, albeit at a higher computational load (the current OF computation from a pair of frames from urban image sequence takes 1120 seconds on a dual core processor). A fast computation of flow is feasible with H but with a little trade off in accuracy. While high computational cost may seem prohibitive for real time applications, the methods strength in terms of robustness to distortions and ability to extract structure from motion even from noisy data is of interest, especially in applications such as echocardiography.

Chapter 5

APPLICATIONS OF INFORMATION FLOW

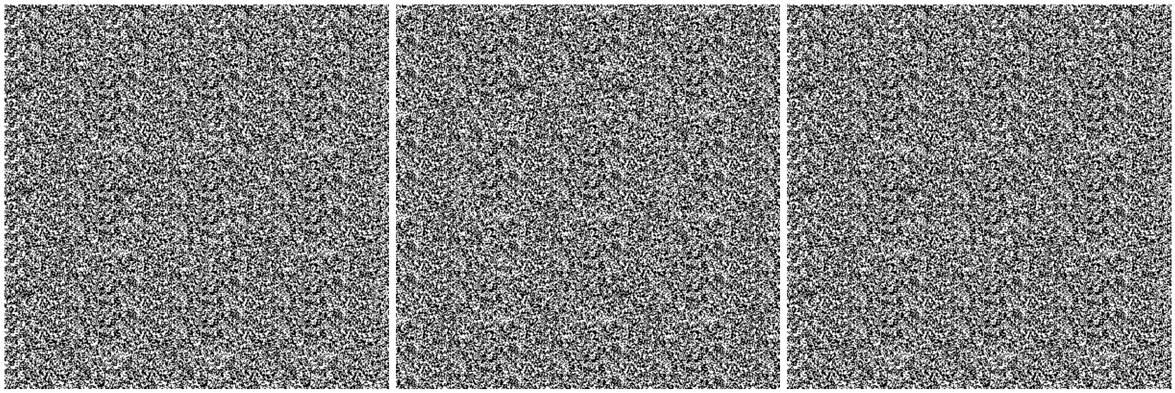
In this chapter, we show that the proposed robust OF formulation can be used in developing systems for applications in both noisy and non-noisy environments. In this chapter, we address three applications of information flow formulation: (i) Motion edge detection and (ii) Heart-rate estimation from Ultrasonography and (iii) Color flow. The conventional OF methods which assume BC may fail in these scenarios in the first two applications. Hence, we have used the proposed *Information flow* formulation for robust and accurate OF estimation.

5.1 Motion edge detection

A well known phenomenon in visual perception is the type II motion as mentioned in Chapter 3. Examples of this type of motion can be found in [2]. Motion edges are edges perceived by humans when the image contains regions filled with coherently moving microtextures. These motion edges in turn, define structures which are apparently in motion these figures. These type of edges can be observed individual frames (see Figure 5.1).

5.1.1 Edges from OF

We tested the proposed OF method to see if a similar structure can be extracted from a computed OF field. Figure 5.2 shows two samples. In the first of these examples, humans perceive a central annulus framed by a rectangle on the periphery, moving upwards against a textured field. The OF field computed by the Classic+NL method and the our method are shown in Figure 5.2. We did some post-processing with 1-D median filters and intensity profiles to get sharp edges. We have scanned the magnitude image (of flow fields) with 1-D median filters along rows and columns individually. This will reduce the fuzzyness at edges. In the second the number of structures increases to three. The OF field corresponding to the Classic+NL is incoherent with the central annulus shape barely visible. In contrast, the outputs of our method show a clear separation of the flow fields into distinct classes. The



(a) frame1 (b) frame2 (c) frame3

Figure 5.1 Some frames from a motion edge sequence

extraction of motion edges and hence the annulus and rectangular frames is quite straightforward from these OF fields.

We can attempt the following explanation for the two surprising percepts seen in Figure 5.2. The two regions seen in the figure are filled with different fields of coherent motion, one horizontal, the other vertical. Thus one will be a bright region in the population of cells sensitive to rightward horizontal motion ('horizontal motion image'), while the other will be bright in the 'vertical motion image'. The edge separating the two regions will be present in both, as and edge between a light and a dark region. In both these images the edge present will carry the corresponding motion signal, much as it might at the boundary between a coherently moving textured area and a dark background. Thus the horizontal motion image will generate a horizontally moving edge percept, and the vertical motion image will generate a vertically moving edge percept. One of these two will repress the other and thus 'capture' the edge, much as presumably happens in binocular fusion. If the edge is captured by the vertically moving percept, the region outside the circle will be seen as having an edge moving coherently downward and so will be perceived as an area with a circular hole in it moving downward. If the edge is captured by the horizontally moving percept, the circle will have an edge moving horizontally and so will be perceived as an object moving horizontally.

There is much interest in explaining the coherence of such fundamental gestalts as edges. So, it is reasonable to detect those motion edges in a given random dot image sequence. These data sequences are downloaded from [2]. We locate these gestalts by applying our method (with NKD criterion) on any two adjacent frames of given image sequence and then finding edges on color maps of flow. We observed that our method is unable to find correct edge corners. However, this issue can be solved by using some post-processing steps. Final Edge detection results are as shown in Figure 5.2.

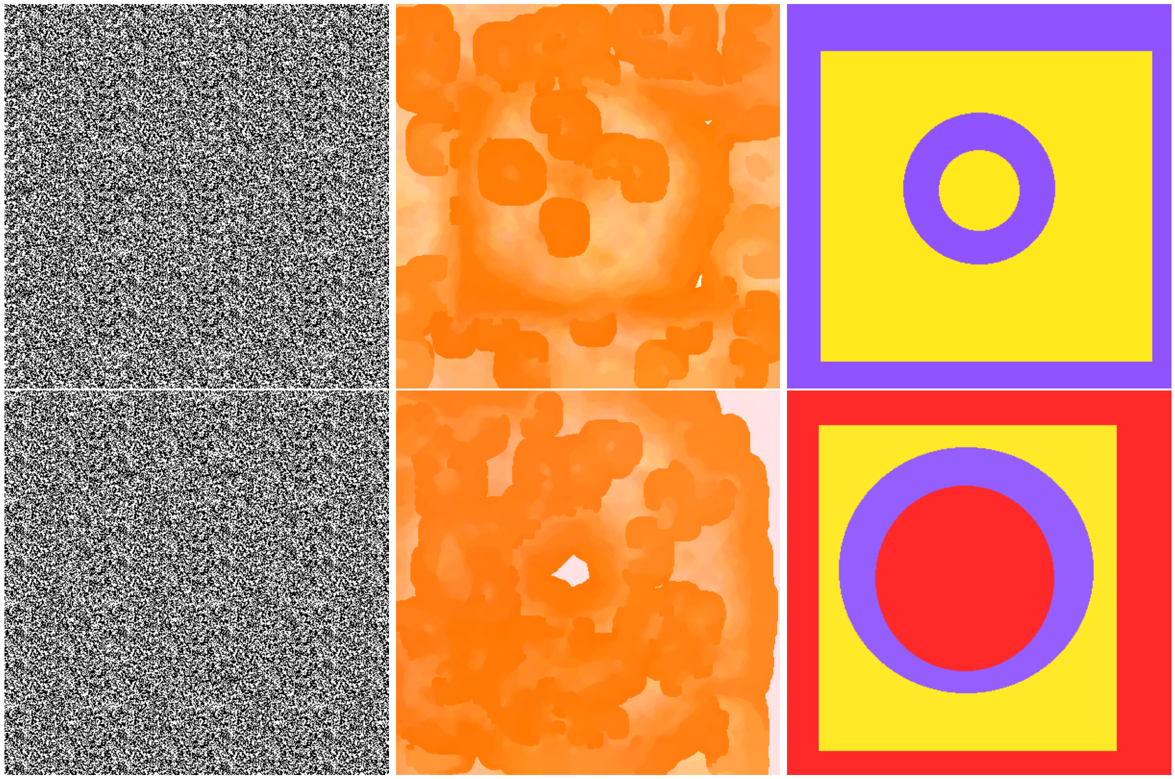


Figure 5.2 Detection of motion edges by using our method: Left column: a frame from 2 input image sequences; color coded flow fields of Classic+NL method [80] (middle column:) and the proposed method (right column).

5.2 Heart-rate estimation from Ultrasonography

Heart-rate (HR) is very important parameter in analysing the vascular blood supply and in the identification of pathological abnormalities. In general, physician records ECG also while cardiac image acquisition to know HR and position of beats [43]. When comparing cardiac motion in Ultrasound (US) sequence and ECG, it should be kept in mind that mechanical and electrical activity are involved and are not identical. Although both quantities arise from the cardiac cycle, there is a variation of the time interval between electrical activation and the main ejection of blood of approximately eight milliseconds [82]. Hence, using ECG signal to know the HR beats in US image sequence is not right strategy. Hence, there is a clear need to develop algorithms which estimate the HR and length of each cardiac cycle (beat-to-beat resolution) [16]. Therefore, we track some myocardial landmarks temporally to estimate HR and beats.

Underlying the heartbeat is the myocardial cyclic motion. Hence, a system for estimating the Heartbeat Rate (HR) can be designed as follows: 1. Detection of landmarks (points on myocardial boundary);

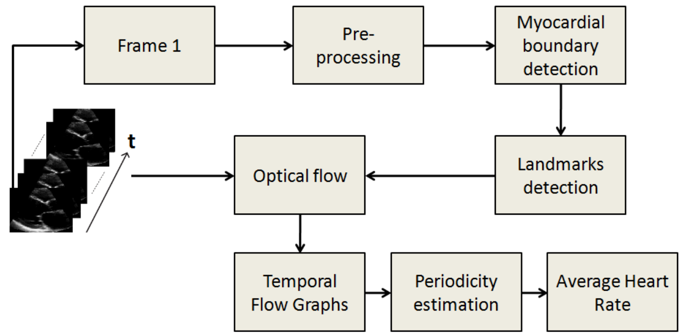


Figure 5.3 Flow chart of the HR estimation

2. Tracking of landmarks using Optical flow and 3. Average HR estimation. The flow chart of HR estimation system is as shown in Figure 5.6.

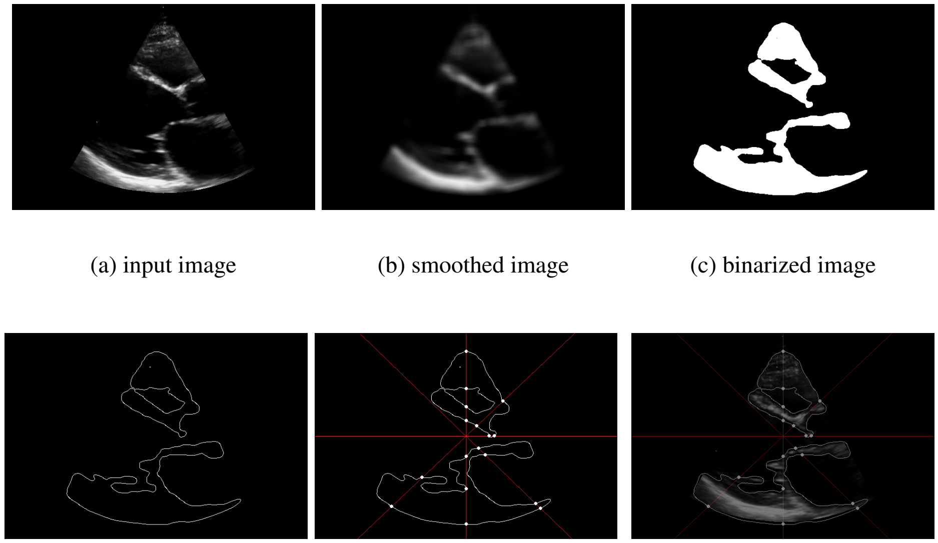
5.2.1 Landmarks detection

Points on myocardial boundary are of interest as they follows periodic motion rather than the points on inner part of the heart. Automatic detection of Myocardial boundary is a difficult task because of the speckle noise present in US images. To reduce the influence of speckle noise, the original image is filtered by an averaging mask of size 15x15. Figure 5.4(b) shows the result of applying the filtering process to the echocardiographic images shown in Figure 5.4(a).
Morphological Operations: The smoothed image is converted to a binary image using a thresholding operation. The thereshold can be found empirically determined. The resulting binary image may contain several small holes in the middle of posterior segment of left ventricle, generated by speckle noise or artefacts that are not previously eliminated by smoothing. These holes are filled using a morphological closing with structuring element of 5x5 to reduce unwanted or noisy segments. The desired myocardial boundary (Figure 5.4(d)) is the inside edge of the binary image. This is found using a Canny edge detector.
Points detection : Any point on myocardial boundary is suitable as point of interest as it follows periodic path while cardiac motion. So, we selected a point set (after rejection of points on/nearer to the image boundary) which are the intersection points of walls map with the fundamental lines x=0, y=0, x=y and x=-y (see Figure 5.4(e)).

5.2.2 Tracking of landmarks

To track or analyse the cyclic motion of landmarks, we defined a signal called Temporal Flow Graph (TFG) as follows.

Consider a point of interest undergoing motion captured in an video sequence. The displacement/motion of this point over two frames is represented by a vector v. Estimation of such a vector at every point



(a) input image (b) smoothed image (c) binarized image

(d) canny edge detection (e) Intersecting points of edge map and lines (f) overlapped with raw image

Figure 5.4 Landmarks detection

over the entire sequence yields a dense motion field commonly referred to as optical flow. Let us consider quantising the direction (or angle, θ) of the motion vector to be +1 when $\theta < 180^0$ and -1 when $\theta > 180^0$. This helps represent the displacement of over time as a 1-D function of time where the value of the function is determined by the displacement magnitude and the sign is determined by the displacement direction. Let I be a given image sequence and let the instantaneous displacement at a point of interest between the n^{th} and $(n+1)^{th}$ frame be of magnitude v_{pn} in the direction θ_{pn} which is quantised as follows

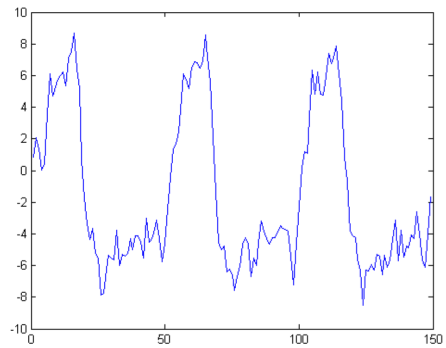
$$dir(\theta_n) = \begin{cases} -1, & -180^0 \leq \theta_n < 0^0 \\ +1, & 0^0 \leq \theta_n < 180^0 \end{cases}$$

The cumulative sum of the instantaneous displacement of a point helps track the motion flow of that point over time. We can view such a sum as a temporal flow graph (TFG).

$$TFG(k) = \sum_{n=1}^{k} v_n.dir(\theta_n) \tag{5.1}$$

where k is the frame index. From Eq.5.1, we can observe that the TFG of a point undergoing cyclic motion will be a periodic function. We can use this fact to analyse the heart-beat rate variability and abnormalities in cyclic motion.

The parameters (v_n, θ_n) in Eq.5.1 were derived using the proposed information flow (OF) algorithm by using the D metric (Eq. 4.2).



(a)

Figure 5.5 The TFG for a myocardial boundary point

The TFG graph (Magnitude of v vs. frame number) of a random point on myocardial boundary is shown in Figure 5.5. Every peak in this graph corresponds to a beat.

Heartbeat estimation: The periodicity (in general quasi-periodicity) of each TFG signal can be estimated using any periodicity estimation method such as [37]. The beat to beat (peak to peak) intervals can be found from the TFG following which the median of the peak intervals yields the desired periodicity of the corresponding landmark. In general, it is better to determine the periodicity (P) of heart motion as the average value estimated from many co-located landmarks. Finally, Average heartbeat rate can be calculated as

$$HR = \frac{P}{fps} \times 60 \qquad (beatsperminute) \tag{5.2}$$

where fps is the frames per second rate of the acquired data.

The proposed HR estimation method was implemented and evaluated over a set of US sequences (recorded on different subjects) sourced from a local hospital. The dataset contained echo videos (of 3 minutes duration each) recorded on 19 patients. The ECG reference was also collected and taken to be the ground-truth. The difference between the HR estimated by our method and the ECG is the error is estimation. These error values are presented in Table 5.1. Included here are results when the optical flow is computed using two other methods.

Table 5.1 Comparison of HR rates

	US_1	US_2	US_3	US_4	US_5	US_6
$Classic + NL_{6.2}$	8.2	9.0	5.3	3.8	-6.6	-4.2
$kNN - ME_{3.0}$	2.1	-3.7	3.4	2.5	-4.3	-1.5
Our method$_{1.7}$	0.4	3.5	-1.6	-1.4	0.7	-2.2

From the Table 5.1, it can be clearly seen that proposed scheme is performing well as the results are nearer to the ECG. For a fair comparison, the optical flow computation using Classic+NL was preceded by a denoising stage as the method is very sensitive to noise. The kNN-ME methods is designed to be robust to noise and hence no denoising was done. Nevertheless, the tabulated results show that the error values tend to be lower with the proposed method for almost all sequences. The first column lists (in the suffix) the mean absolute error for each method across 6 sequences. These values indicate that the proposed method even outperforms the kNN-ME method.

5.3 Color Flow

Colorization is the process of addition of color to a black and white video or still image. A colored image is a vector valued function often represented as three separate channels (Y,C_b and C_r). A gray scale image, in contrast, is a scalar valued function. Thus, the colorization process requires mapping of a scalar to a vector valued function which has no unique solution.

A number of automated and semi-automated techniques exist for colorizing monochrome video [46]-[88]. The prior art systems suffer from one or more of the following disadvantages: (a) slow processing speed, with each frame to be colorized requiring many minutes or hours, resulting in unduly high processing cost [83]; (b) lack of operator convenience and flexibility (selection of seed points/regions on each frame and colors to them) that is needed to facilitate obtainment of quality colorized video [70]; (c) high computational complexity and/or cost(money) of the colorizing system itself [83].

In order to decrease the amount of user interaction, we propose a fully-automated process to colorize a sequence representing a *single scene* which is extendable to a general video sequence. For every single-scene sequence, we assume that there is a frame which contains all the objects. We call this frame as the Most Informative (MI) frame. We propose colorizing this MI frame by using an existing static image colorization method followed by propagation of colors from colorized MI frame to remaining frames based on the motion vectors between frames. Motion estimation is performed using an optical flow technique. All pixels which are not colorized at the end of the previous step is colorized by a refinement step to assign colors. A general video sequence can consist of several scenes. Hence, our colorization strategy requires pre-segmenting the entire video into several scenes using a shot detection method [17], prior to colorization.

5.3.1 Proposed method

Let the given scene/shot be denoted as $I(n)$. The proposed system for colorization of $I(n)$ mainly consists of three parts: (a) MI frame selection; (b) Colorizing the MI frame; (c) Optical flow (OF) computation for all frames by taking MI frame as the starting frame. Thus, if the $n = k$ is the MI frame, then OF is computed between the frame pairs $I(k \pm i)$, $I(k \pm (i + 1))$; $i = 0, 1, 2..$; (d) Propagation of color; (e) Refinement for colorizing the remaining pixels. The block diagram of the proposed algorithm is given in Fig.5.6.

5.3.1.1 Most Informative (MI) frame selection

The MI frame is defined to be the frame which contains maximum number of objects present in that scene. Hence, this frame should have maximum spatial activity (entropy) and highest amount of edges across the sequence. We detected this frame by maximizing a score S which is defined as,

$$S(n) = w_h \frac{H(n)}{\sigma_h} + w_e \frac{E(n)}{\sigma_e} \tag{5.3}$$

Here, σ_h and σ_e are the standard deviations of $H(n)$ and $E(n)$ respectively. w_h and w_e are empirically determined weights. The first term $H(n)$ captures to the spatial activity and the second term $E(n)$ corresponds to the amount of edge content of the n^{th} frame. $H(n)$ is the entropy of the n^{th} frame and is given as

$$H(n) = -\sum_x p(x,n) \log_2(p(x,n)) \tag{5.4}$$

where $p(x,n)$ is the probability of the grey value x in the intensity histogram of the n^{th}-frame. For an 8-bit image, $H(n)\epsilon[0,8]$.
The amount of edge content present in the n^{th}-frame is determined by computing the energy of the gradient of the frame I_n.

$$E(n) = \sum_x \sum_y |\frac{\partial I_n}{\partial x} + \frac{\partial I_n}{\partial y}|^2 \tag{5.5}$$

Finally, the MI frame is selected as the frame with the highest score.

$$k = argmax_{k\epsilon n} S(n) \tag{5.6}$$

where k is the frame number of MI frame.

The above summarized procedure for MI frame can also be used to automatically extract from a video sequence a single key frame representative of its content.

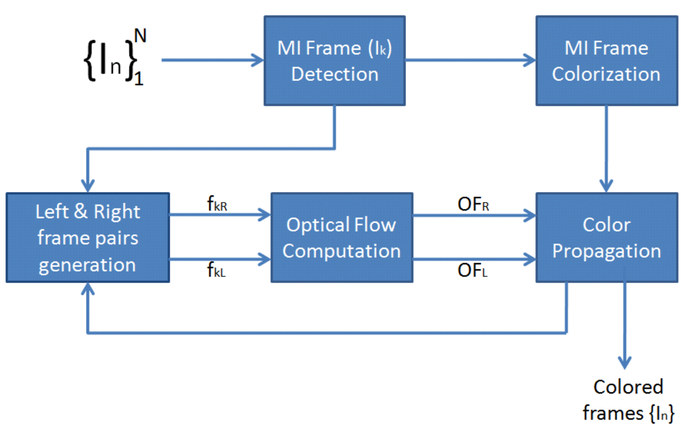


Figure 5.6 Block diagram of the colorization scheme

5.3.1.2 Colorizing the MI frame

After the selection of MI frame, denoted as I_k in Fig.5.6, we colorize the MI frame using a scribble based colorization process from [46] as explained earlier.

The information flow method (with H) was used for motion field estimation to generate the desired motion vectors at every pixel.

5.3.1.3 Color propagation

This is the main task in the colorization process. We assign or propagate the color to a pixel in the current frame from its corresponding location in the neighbouring frame. We do this process starting with the coloured MI frame I_k. Given I_k, the neighbouring frame pairs to its right and left are first determined. These are denoted as $f_{kL} = (I_{k-i}, I_{k-i-1})$ and $f_{kR} = (I_{k+j}, I_{k+j+1})$ respectively. The optical flow is computed for each pair and the color of the k^{th} frame is propagated to the neighbouring frame (I_{k-i} and I_{k+i}) based on the flow information. This is done iteratively until the end of the sequence is reached in both right/left directions. Thus, in the first iteration ($j = 0$), the colors of pixels in I_{k+1} are found from the colors from I_k, according to the motion between these two frames. Likewise, in the next iteration ($j = 1$), the pixels in I_{k+2} will inherit colors from I_{k+1} according to the motion between these frames. This process repeats until the last frame (N). A similar process is followed to color frames which are left neighbours of the MI frame.

5.3.1.4 Refinement

It is possible that some pixels are missed in the process of color propagation between the frames. Colorizing them is the refinement process. This is based on a test for similarity between the greyvalues of the pixel to be colourized (missed pixel) and its neighbours which are already colorized: (1) If the intensity (Y) of the missed pixel is similar to its neighbouring pixel, then they should have same colors (chromatic values: C_b, C_r). (2) If a set of connected pixels are missing and the size of this cluster is more than 5x5, then it either signals the introduction of a new object into the scene or shadow formation due to a change in illumination. This is best resolved by involving the user. Hence, in this case, the user is asked to decide the color with scribble or marker.
The proposed scheme should result in the requirement of a relatively smaller number of user input. This was also seen to be true when experimenting with our large variety of videos.

5.3.2 Results

The proposed colorization system was implemented in MATLAB-R2009a on Windows7 ultimate (32-bit OS), Processor Intel(R)Core2Duo 1.83 GHz, RAM 3GB. It was tested on videos from Levin's data base [46] and animated videos which were independently obtained. In all cases, the system was able to produce high quality colored videos in a short time. We present some sample in this section. In order

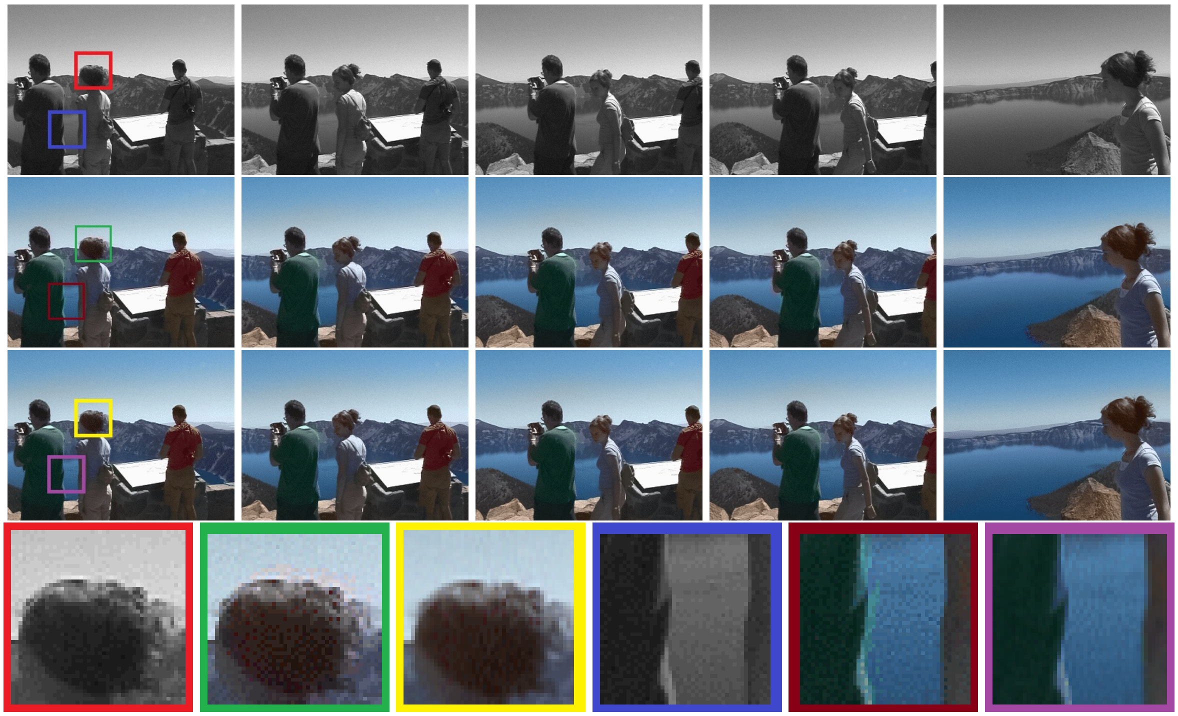


Figure 5.7 Colorization of an Lake scene from a 83-frame clip. First row: greyscale input frames (1,9,18 (MI frame),67,83); Corresponding frames colorized by the proposed (second row) and scribble based methods [46] (third row); Sample zoomed regions are shown in the bottom row.

to assess the performance of the proposed method a set of comparisons were carried out: Comparison with (i) state of art, (ii) ground truth and (iii) some challenging sequences like animation sequences.

5.3.2.1 Comparison with State of Art

Fig.5.7 shows some selected frames from a greyscale movie clip (containing 83 frames) and the corresponding colorized frames. Our system detected frame 18 as the MI frame and did not ask for user interaction for the entire movie clip except for MI frame colorization. The colorization process of this video took about 1 minute and 42 seconds. For comparison, results of colorization with the method in [46] is also shown. This method required 12 scribbles from the user. Our results indicate comparable quality of results. Since our method uses Optical Flow only to define the local temporal neighborhood, it is robust to tracking failures. Some zoomed details are also provided in the bottom row for a detailed comparison from which we can observe a small amount of overlapping between colors at the edges. This is because of the erroneous motion vectors given by OF method at edges. However, these will be invisible while playing a movie.

Sample MI (frame 27) and some of preceding (frame 5) and later (frame 33) frames for a scene captured by a camera in a car are provided in Fig.5.8. The reflection on the windscreen has a subtle change in colour which is successfully propagated in the distant (5^{th}) as well as relatively proximal (33^{rd})

frames. The effectiveness of the proposed system and color propagation are better observed in videos which have been made available at `http://web.iiit.ac.in/~vsrao/colorization`.

(a) 5 (b) 27 (c) 33

Figure 5.8 Colorization of a car video scene. From left to right: 5^{th}, 27^{th} (MI frame) and 33^{rd} frames. Top row: greyscale input; Bottom row: colorized result.

5.3.2.2 Comparison with Ground-truth

We also evaluated our system on colour videos by colorising the greyscale version and using the original colour as ground truth for validation. Some sample frames of a party scene (consisting of 63 frames) are shown in Fig.5.9. For this scene, frame 19 was detected as the MI frame. No other user input was given. The results for the distal (5^{th} and (45^{th}) frames still appear to be very close in quality to the original colour frames. This is also illustrated by using $PSNR$ measure to quantify error in colourisation. The PSNR value for the n^{th} colorized frame is given by,

$$PSNR(n) = 20\log_{10}(\frac{255}{MSE(n)}) \tag{5.7}$$

where MSE(n) is the mean squared error between the original and the colorized n^{th}frame. Fig.5.10 shows the PSNR plot for 63 frames, with the x-axis representing the frame number. Generally, the higher the PSNR value the more similar is the colored image to the original one. The plot peaks at 19^{th} frame which is the MI frame. The first and final frames have least PSNR. The PSNR degrades for non-MI frames due to the color propagation error. However, given the highly magnified scale for the y-axis

Figure 5.9 Colorization of a party video scene. From left to right: 5^{th}, 19^{th} (MI frame) and 45^{th} frames. First row: greyscale input; second row: colorized result; third row: ground truth.

in this plot, the degradation is only by 2 $\%$ which demonstrates the effectiveness of the propagation strategy. The entire colorization process of 63 frames took 93 seconds.

5.3.2.3 Animated movie scenes

Colorisation on animated movie scenes from *Finding Nemo* and *Megamind* were also tested. Fig.5.11 shows our result and ground truth (GT) version of frame 31 from a scene of *Finding Nemo* movie. We can clearly observe that there is very little overlap between color content of objects at edges. In general, animated movie scenes contain large motion fields. Hence, optical flow algorithms might be more erroneous in such cases. Large displacement optical flow methods are more suitable for such sequences, however they are computationally expensive. Sudden appearance of objects in a scene is another characteristic of such sequences. Accordingly, the required number of user interaction can increase for our system. In our experiments with a large dataset, the maximum number of times that user interacted to decide color with scribbles was found to be 16 for a clip of 274 frames from *Megamind* animated movie.

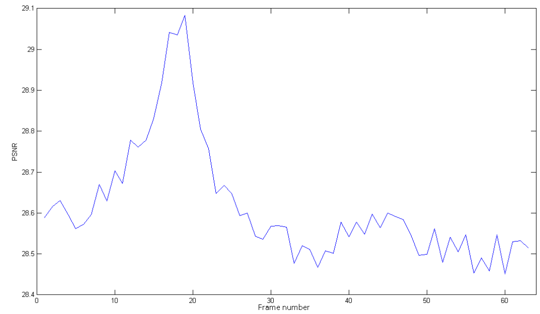


Figure 5.10 PSNR plot for the party movie scene with 63 frames.

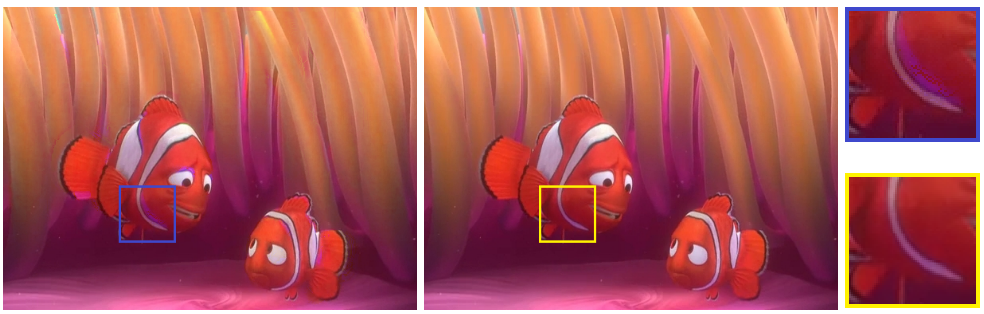


Figure 5.11 Ground truth and colorized result for frame 31 of *Finding Nemo* video scene. (a) our result, (b) original frame (GT), (c) zoomed regions.

5.3.2.4 Robustness

In real scenarios, there may be videos with noise and/or low resolution (surveillance for instance). Classical approaches result in erroneous OF in such scenarios which in turn results in accurate color propagation. We have tested the performance of the *information flow* based approach on a toddler image sequence which is of low resolution (180x240). The key benefit, robustness of the method was tested by imposing noisy environment (adding salt&pepper noise) to *toddler* image sequence. Frame36 and its version which contains 15% noise and their corresponding colorized frames are shown in Fig.5.12. From the figure, it can be observed that salt & pepper noise does not affect the color propagation framework and on-off points (due to salt & pepper noise) also got colorized with neighboring colors. The former one due to robust nature of the *information flow* formulation and the later one due to refinement step in color propagation method.

Run time Statistics: Key features of the proposed system are computational simplicity and a greatly reduced need for user interaction. This is demonstrated by the time required for computations in our method with that of Levin's method [46] to colorize the scene in Fig.5.8 which has 60 frames. [46] reports that the user interaction is used for the first frame and 11 other frames. The required computations are listed in Table 1. Here, U is the unit time taken by the user to specify colors to seed points in a frame which is a minimum of 30 seconds. F is the unit time for colorization of the MI frame and M1, M2 are the unit times taken for color propagation between two frames by [46] and our methods respectively.

Figure 5.12 Color propagation in Noisy environments: (a)Frame36; (b) Noisy frame36; (c)colorized frame of (a); and (d) colorized frame of (b).

Task	Method[46]	Our method
MI frame selection	0	0.8 sec
Colorizing key frame	U+F	U+F
User interactions	11(U+F)	0
Color propagation	48M1	59M2
Total Time	12U+12F+48M1 $\sim$ 17.04 min	U+F+0.8+59M2 $\sim$ 2.33 min

M2 is inclusive of the refinement time. The total time shown for [46] is the time taken by using the code available at `http://www.cs.huji.ac.il/~yweiss/Colorization/`.

5.4 Summary

In this chapter, we have looked at three applications which uses information flow framework for robust motion field estimation. Those applications include motion edge detection, HR estimation in ultrasonography and Color propagation in film colorization.

Chapter 6

MOTION BLUR

In this chapter, we discuss about the motion blur which is a common problem in capturing videos with a fast moving camera/objects. We consider deblurring the frames which are blurred due to fast moving camera. In general, fast moving camera (not rotating) produces uniform blur in frames. We propose a method to deblur these kind of images by using statistical information only. In this work, we start with a hypothesis that there is sufficient information within a blurred image and approach the deblurring problem as an optimisation process where the de-blurring is to be done by satisfying a set of conditions which includes constraints on statistical information measures: Shannon's entropy and Burg's entropy. These conditions are derived from first principles underlying the degradation process assuming noise-free environments. We propose a novel but effective method for removing motion blur from a single blurred image via an iterative algorithm. The strength of this method is that it enables deblurring without resorting to estimation of the blur kernel or blur depth. The proposed iterative method has been tested on several images with different degrees of blur. The obtained results have been compared with state of the art techniques including those that require more than one input image. The results are consistently of high quality and comparable or superior to the existing methods which demonstrates the effectiveness of the proposed technique.

6.1 Introduction

Very often images are corrupted by motion blur due to the relative motion between the camera and the scene during the integration time of the image. Recovering un-blurred image from a motion blurred image has long been a fundamental research problem in digital imaging. The standard way to express the relationship between the observed image $g(i,j)$ and its uncorrupted version $f(i,j)$ in noise-free environments is

$$g(i,j) = f(i,j) \; * \; h(i,j) \tag{6.1}$$

where h is the blur kernel or point spread function (PSF) and $*$ is the convolution operator. Numerous methods have been proposed in the past for motion de-blurring. If one assumes that the blur kernel is shift-invariant, the problem reduces to that of image de-convolution. Image de-convolution can be further separated into the blind and non-blind cases. In non-blind de-convolution, the motion blur kernel is assumed to be known or computed elsewhere; the only task remaining is to estimate the un-blurred latent image. Traditional methods such as Weiner filtering and Richardson-Lucy (RL) de-convolution [51] were proposed decades ago, but continue to be widely used in many image restoration tasks because they are simple and efficient. However, these methods tend to suffer from unpleasant ringing artifacts that appear near strong edges. In the case of blind de-convolution [24] [38], the problem is even more ill-posed, since both the blur kernel and latent image are assumed unknown. The complexity of natural image structures and diversity of blur kernel shapes make it easy to over- or under-fit probabilistic priors [24].

In this work, we begin our investigation of the blind de-convolution problem by exploring the major causes of visual artifacts such as ringing. Our study shows that the performance of current de-convolution methods is highly dependent on accurate estimation of motion blur parameters. We therefore observe that a better model of de-blurring and a more explicit handling of visual artifacts caused by the blur kernel estimate errors should substantially improve results. Based on these ideas, we propose an approach in which deblurring is achieved iteratively without explicitly estimating the blur kernel, by satisfying a set of conditions.

6.2 Related Work

We first review techniques for non-blind de-convolution, where the blur kernel is known and only a latent image must be recovered from the observed, blurred image. The most common technique is the RL technique for de-convolution [51], which computes the latent image with the assumption that its pixel intensities conform to a Poisson distribution. Donatelli et al. [21] use a PDE-based model to recover a latent image with reduced ringing by incorporating an anti-reflective boundary condition and a re-blurring step. A common approach in the signal processing community to the de-convolution problem is to transpose the problem to the wavelet or the frequency domain (an example is [62]); However, many of these papers lack experiments in de-blurring real photographs, and few of them attempt to model error in the estimated kernel. Levin et al. [45] use a sparse derivative prior to avoid ringing artifacts in de-convolution. Most non-blind de-convolution methods assume that the blur kernel contains no errors, however, even small kernel errors can lead to significant artifacts. Finally, many of these de-convolution methods require complex parameter settings and long computation times.

Blind de-convolution is a significantly more challenging and ill-posed problem, since the blur kernel is also unknown. Some techniques make the problem more tractable by leveraging additional input, such as multiple images. Rav-Acha et al. [68] utilise the information in two motion blurred images, while Yuan et al. [89] use a pair of images, one blurred and one noisy, to facilitate capture in low light conditions. Another strategy adopted has been to take advantage of additional, specialized hardware. Ben-Ezra and Nayar [8] attach a low-resolution video camera to a high-resolution still camera to help in recording the blur kernel. Raskar et al. [67] flutter the opening and closing of the camera shutter during exposure to minimize the loss of high spatial frequencies. This method requires the object motion path to be specified by the user. The most ill-posed problem is single-image blind de-convolution, which must both estimate the PSF and the latent image. Early approaches usually assume simple parametric models for the PSF such as a low-pass filter in the frequency domain [42] or a sum of normal distributions [47]. Fergus et al.[24] showed that blur kernels are often complex and sharp; they use ensemble learning (Miskin and MacKay [58]) to recover a blur kernel while assuming a some statistical distribution for natural image gradients. A variational method is used to approximate the posterior distribution and the RL technique is used for de-convolution. Jia et al. [38] recovered the PSF from the perspective of transparency by assuming the transparency map of a clear foreground object should be two-tone. This method is limited by a need to find regions that produce high quality matting results. Qi shan et al. [72] creates an unified probabilistic framework for both blur kernel estimation and latent image recovery by allowing these two estimation problems to interact to avoid local minima and ringing artifacts.

Our hypothesis is that there is sufficient information in the blurred image to aid deblurring process. Accordingly we aim to devise a solution which takes a novel different approach to the problem. We first present the necessary basics and then present the proposed method.

6.3 Modeling Motion Blur

Let us assume that a linear, non-recursive (FIR) model represents the degradation of digital (sampled) images caused by motion blur. The original, blur free $M \times N$ image f is convolved with a blur kernel h. De-blurring images requires the application of the de-blurring operator D, which produces a de-blurred image f when applied to the blurred image g, that is $D(g) = f$.

The blur kernel provides information of the underlying motion during the capture process. In the most simple case, such as for a uniform linear motion along the x-axis with a speed of k pixels during the capturing period, the PSF is given by a one-dimensional vector of the length k+1:

$$h_{lin} = \frac{1}{k+1}[111\dots1] \tag{6.2}$$

[8] propose a method to determine the motion paths during the capturing process. Their analysis shows that the model for the PSF has to be extended to represent motion in a two-dimensional plane. The PSF is a matrix h of size $U \times V$, where each entry $h(i,j)$ i=$1,2,\dots,U$, j=$1,2,\dots,V$ represents the percentage the camera has been displaced by $i-(U/2), j-(V/2)$ from the centre during the capture.

$$h = \frac{1}{K}\begin{bmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,V} \\ h_{2,1} & h_{2,2} & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ h_{U,1} & h_{U,2} & \dots & h_{U,V} \end{bmatrix} \tag{6.3}$$

Where the parameter K is a normalizing constant to ensure that the sum over the entries of the matrix equals to 1.

6.4 Proposed Method

The proposed method consists of two parts. i) direction detection to estimate the direction of motion (ϕ) and ii) compensation for blur. These are presented in detail below.

6.4.1 Direction Detection

Since motion blur is essentially directional averaging, it results in parallel white bands in the Fourier spectrum of a degraded image. This has been used to effectively determine the direction (ϕ) of the motion blur [60] [61]. We extract the blur direction using the same principle but using the Radon transform: Let $|G(u, v)|$ be the amplitude spectrum of the given blurred image $g[m, n]$. We take the Radon transform (RT) of this function $|G(u, v)|$ to find the direction of these bands and find the angle corresponding to the maxima in the RT. After finding the motion direction estimation, the blurred image g is rotated to align it with the computed motion direction. The desired deblurred image is estimated by compensating for the blur as described next.

6.4.2 Compensation

Deblurring can be viewed as a problem where a set of corrupted data (blurred pixel values) is given and the process of deblurring has to recover the original pixel values while satisfying some conditions. This leads to casting the compensation step as an optimisation process which satisfies a set of conditions. The requisite conditions can be identified from the basic principles underlying the blur process.

6.4.2.1 C1. Conservation of Mass

If the blur kernel is a normalized one, the mean value of the signal will not change after convolution. Given that the blur kernel in eq 6.3 is normalized, this implies the sum of all pixel values in the blurred image must equal to that in the restored image [20]. The sum of all pixel values in blurred image as M_1 is given as

$$M_1 = \sum \sum g(i, j) \tag{6.4}$$

6.4.2.2 C2. Conservation of Energy

The degradation process obeys Law of conservation of energy as the motion of an object or of the camera does not need any optical energy [20]. Hence, the energy of a blurred image is same as that in the original image. This energy denoted by M_2 is

$$M_2 = \sum \sum g(i, j)^2 \tag{6.5}$$

6.4.2.3 C3. Entropy condition

Many restoration algorithms are based on minimization of Shannons entropy E (examples are [20],[63]), which is given as

$$E = -\sum\sum g(i,j) \;\; log[g(i,j)] \tag{6.6}$$

The basic assumption behind these methods is that the Shannons entropy of the original image is less than that of the degraded image. This may not hold for very large-size blur kernels. We have found that the entropy of images increases with blur depth up to a certain level, after which it starts decreasing. Hence, we include the next condition.

6.4.2.4 C4. Information condition

For a related inversion problem in speech processing, an alternate measure for entropy, namely the Burg entropy is used which is defined as

$$B = -\sum\sum log[g(i,j)] \tag{6.7}$$

Burg's entropy has been argued to be a better representation of information content and has previously been used in image reconstruction [4]. In the context of restoration, it has been shown that B value of a restored image is higher than that of the corrupted source image [63]. In the proposed method, this entropy measure is used and deblurring aims to maximise the same.

6.4.2.5 Compensate Function

Given a current pixel value in a motion blurred image, its value is likely to be due to an averaging process over its immediate neighbours. Hence, a compensate function C_f is defined to reverse this process. The function for two adjacent pixels is defined as follows:

$$C_f^2(i,j) = a.g(i,j) - b.g(i,j-1) - c.g(i,j+1) \tag{6.8}$$

The function for four adjacent pixels is defined as

$$\begin{aligned} C_f^4(i,j) = a.g(i,j) - b.g(i,j-1) - c.g(i,j+1) \\ -d.g(i,j-2) - e.g(i,j+2) \end{aligned} \tag{6.9}$$

where a, b, c, d and e are unknown re-weight parameters which will be found iteratively. An illustration for processing a row of pixels is shown in Figure 6.1.

From the Figure 6.1, it can be seen that estimation of a current pixel depends on 3 pixels from the previous iteration. Hence, after k iterations, estimation of a current pixel depends on 3k pixels in the input blurred image. So the number required iterations is indirectly based on the length of blur (L). In each iteration, the optimum values of of the re-weight parameters are estimated by imposing the condition set C1 through C4. Next, we present an algorithm for the same. For simplicity we assume a C_f^2 case.

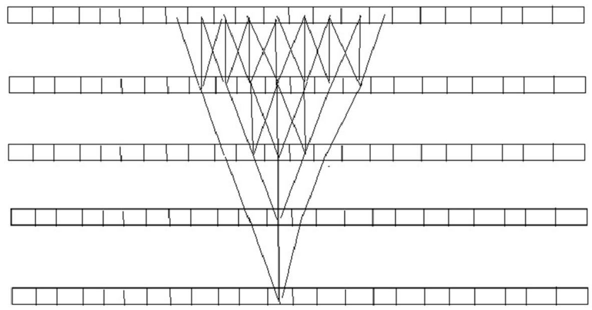


Figure 6.1 Four iterations of a row in an image using C_f^2.

6.4.2.6 Algorithm for Optimization

The problem at hand is optimization of weight parameters a, b, and c with respect to the condition set. The uni-variate method is adopted for a solution of this problem, by multiplying the step size S_i by very small increment ε. In this method, only one parameter is changed at a time to produce a sequence of improved approximations to reach the optimum point. Starting at a base point $P_i = (a, b, c)_i$ in the i^{th} iteration, the value of any one of $(n-1)$ parameters is fixed while others are varied. The purpose is to produce a new base point P_{i+1}. The search is now continued in a new direction. The new direction is obtained by changing any one of the $n-1$ parameters that has been fixed in the previous iteration. After all the n directions are searched sequentially, the first cycle is completed and values of a,b and c are obtained. These are placed in a dummy image which forms the input for the next iteration. The entire process of sequential optimization is repeated until the values of (a,b,c) is approximately (1,0,0). The choice of the direction and the step length in the modified uni-variate method is summarized here.

MODIFIED UNIVARIATE ALGORITHM

1. Choose a starting point P_i = $(a, b, c)_i$ and set i = 1.

2. Find the search direction S_i as 6.8

$$\mathrm{S}_i^T = \begin{cases} ((1,0,0,0,0,\dots) & \text{i=1, n+1, 2n+1} \\ (0,1,0,0,0,\dots) & \text{i=2, n+2, 2n+2} \\ \vdots & \\ (0,0,0,0,0,\dots,1) & \text{i=n, 2n, 3n, \dots} \end{cases}$$

3. For the current direction S_i , find the values of M_1, M_2, E and B and check if condition set is satisfied. If condition set is not satisfied, find whether the entropy (E) values decreases in the positive or negative direction. For this, we take a small probe length (ε), also called learning factor and evaluate $E_i = E(P_i)$, $E_i^+ = E(P_i + \varepsilon S_i)$ and $E_i- = E(P_i - \varepsilon S_i)$. If $E_i^+ > E_i^-$, S_i will be the correct direction for decreasing the values of E_i, and if $E_i^+ < E_i^-$, -S_i will be the correct direction. If both E_i^+ and $E_i^-$are less than E_i , we take P_i as the minimum of the two.

4. Set $P_{i+1} = P_i + \varepsilon S_i$.

(a) (b)

Figure 6.2 Blurred images captured by a hand-held camera and corresponding outputs of our method.

5. $E_i + 1 = E(P_{i+1})$.

6. Set i = i+1 and go to step 2. Continue this procedure until (a,b,c) satisfies the condition set.

We have taken a unit step length for computational simplicity. The algorithm for the de-blurring technique is as follows.

ALGORITHM FOR ITERATIVE MOTION DEBLURRING (IMD)

1. Find the angle of direction of motion (ϕ)

2. Rotate the coordinate system by an angle ϕ.

3. Apply the compensate function to rotated R, G, and B planes of blurred image individually.

4. Impose the condition set using Modified Uni-variate method for each plane.

5. Create dummy image planes with a,b, and c.

6. Repeat 3 to 6 steps with these dummy image planes until we get a=1, b=0, c=0 approximately for each plane.

7. Anti-rotate the image.

8. Display the restored image.

Any algorithm that performs de-convolution in the Fourier domain needs a post processing step to suppress ringing artifacts at the image boundaries; for example, Fergus et al. [24] process the image near boundaries using the Matlab edgetaper command. We instead use the approach of Liu and Jia [49] to suppress the ringing. Some results of this method are provided in Figure 6.2.

6.5 Experimental Results

The proposed iterative deblurring algorithm was tested on numerous images. We present some sample results in this section. Two blurred test images captured using a handheld camera and the corresponding deblurred results obtained by the proposed method is shown in Figure 6.2.

Figure 6.3 (a) motion blurred image used in [68]. (b) Deblurred result from [68] using information from two blurred images. (c) IMD result using only blurred image shown in (a).

In order to assess the performance of to proposed method against existing methods a set of comparisons were carried out: Deblurring i) without use of additional images and ii) with use of additional information/images. Henceforth, the proposed technique is referred to as IMD for convenience.

Deblurring without use of additional images

A uniformly blurred image and the deblurred results are shown in Figure. 6.4(a). The results of RL, Levin et al. [45] and IMD techniques are shown in Figure 6.4 (b), (c) respectively. IMD result exhibits sharper image details and fewer artifacts such as ringing around sharp edges, than the others.

We next illustrate blind de-convolution on two test images taken from [72]; These are shown in Figure 6.5 and Figure 6.6. The blur is due to camera shake. The results of two sample techniques

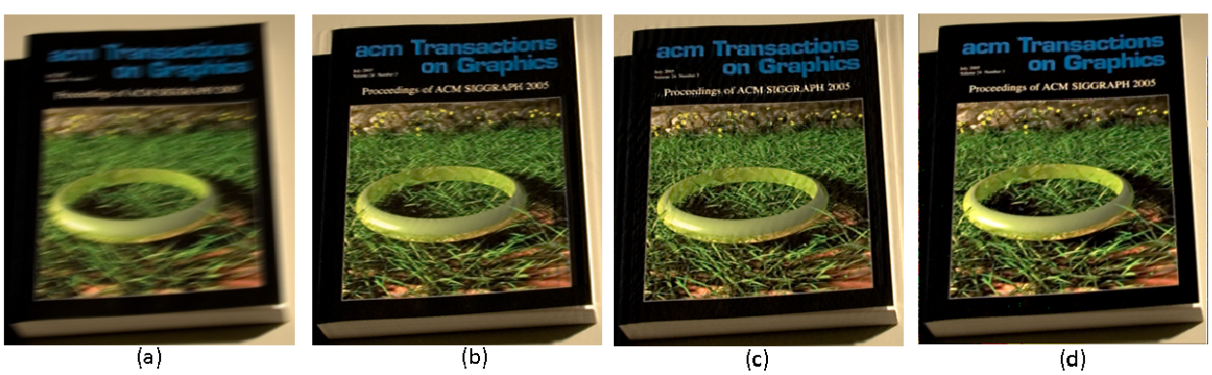


Figure 6.4 Non-blind de-convolution example. (a) blurred image used in [72]. Deblurred results of (b) RL algorithm (c) sparse prior method [45] and (d) IMD.

namely [24] and [38] which are based on the RL technique are also taken from [72]. The degree of blur in the second image shown in Figure 6.6 (a)) is caused by a large-size kernel, which is challenging for kernel estimation. The results of IMD is shown alongside for comparison for both test images. The IMD results for the green toy image is comparable with some areas such as the right ear, being restored better. The colour and sheen are superior in the result of [38]. The IMD results for the second test image in Figure 6.6 (a)) is in comparison clearer compared to the other two techniques. This implies that IMD is superior at handling high degree of blur. A comparison with the most recent deblurring method [72] which uses a probabilistic approach is shown in Figure 6.7. The two results appear to be of similar quality.

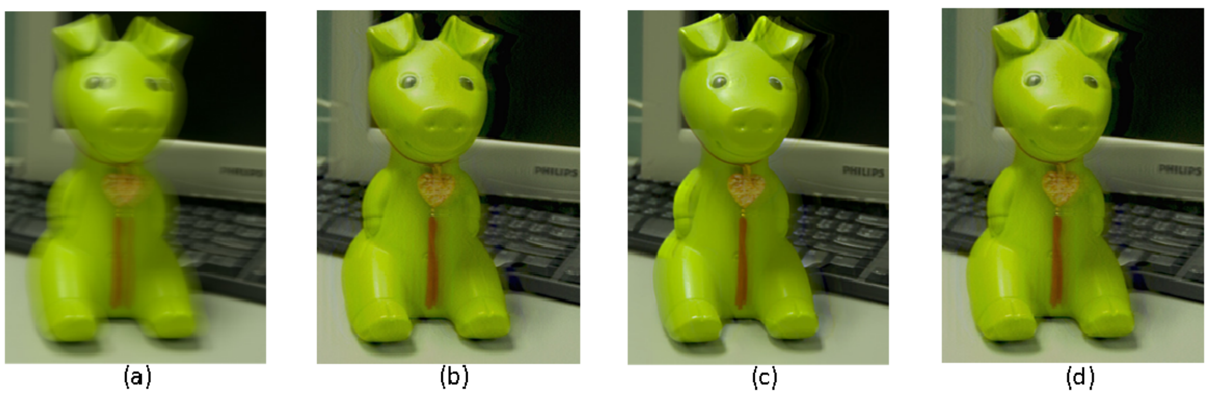


Figure 6.5 Blind deconvolution example 1. a) input blurred images; de-blurring results of b) Fergus et al. [24], c) Jia et al. [38] and d) IMD.

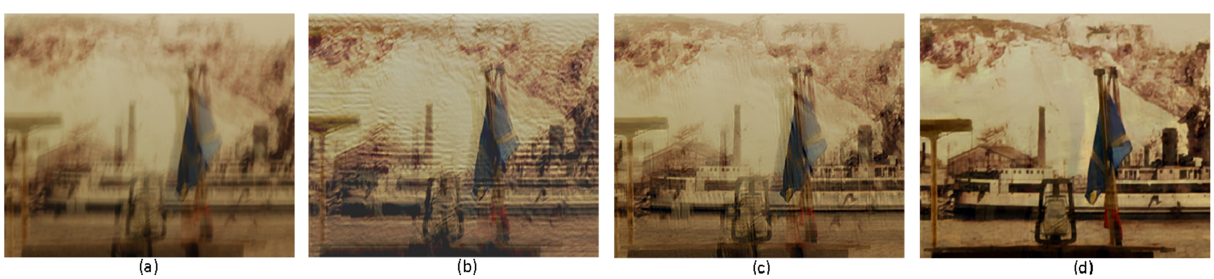


Figure 6.6 Blind deconvolution example 2. a) input blurred image; de-blurring results of b) Fergus et al. [24], c) Jia et al. [38] and d) IMD. Other two methods use RL de-convolution to restore the blurred image.

Deblurring with the use of additional images

In this section, we compare the IMD performance against methods which use additional input. Two blurred images with different camera motions are used in [68] to create the results in Figure 6.3. In comparison, the IMD result based on the first blurred input is remarkably of the same quality. The technique in [89] uses information from two images, one blurred and one noisy, to create the result in Figure 6.8 and Figure. 6.9. Finally, Ben-Ezra and Nayar [8] acquire a blur kernel using a video camera

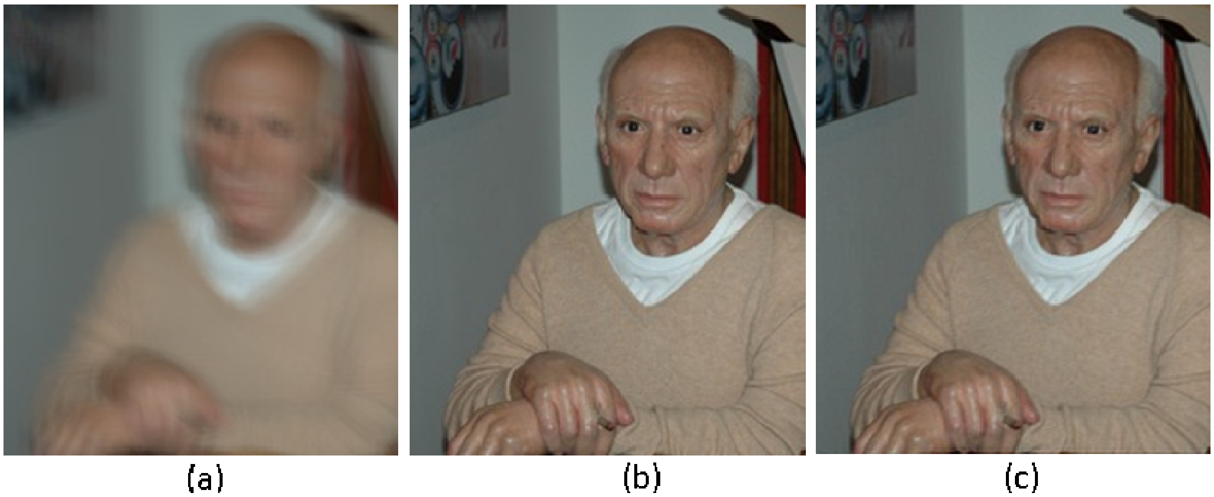


Figure 6.7 Blind deconvolution example 3. a) input image; de-blurring results of b) [72], b) and c) IMD.

Figure 6.8 Deblurring with additional input images from [89]. a) The blurred input image, b) result from [89], c) IMD result with only blurred image as input and d) some close-ups of our results.

that is attached to a still camera, and then use the kernel to deconvolve the blurred photo produced by the still camera. Their result is shown in Figure. 6.10. In comparison with all these three cases, IMD remarkably produces comparable results with just one input image.

Finally, two more challenging real examples and IMD results are shown in Figure 6.11, all containing complex structures and blur from a variety of camera motions. The ringing, even around strong edges and textures, are significantly reduced. The remaining artifact is caused mainly by the fact that the motion blur is not absolutely spatially invariant. Using a hand-held camera, slight camera rotation and motion parallax are easily introduced by Shan et al. [73] .

Figure 6.9 Deblurring with additional input images from [89]. a) The blurred input image, b) result from [89], c) IMD result with only blurred image as input and d) some close-ups of our results.

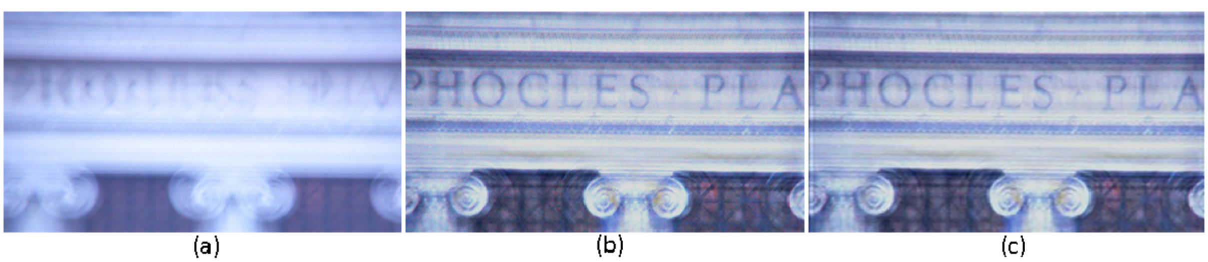


Figure 6.10 Deblurring with additional input images from [8] a) a motion blurred image of a building from the paper of Ben-Ezra and Nayar[8], b) their result using information from an attached video camera to estimate camera motion and c) IMD result obtained with one input image.

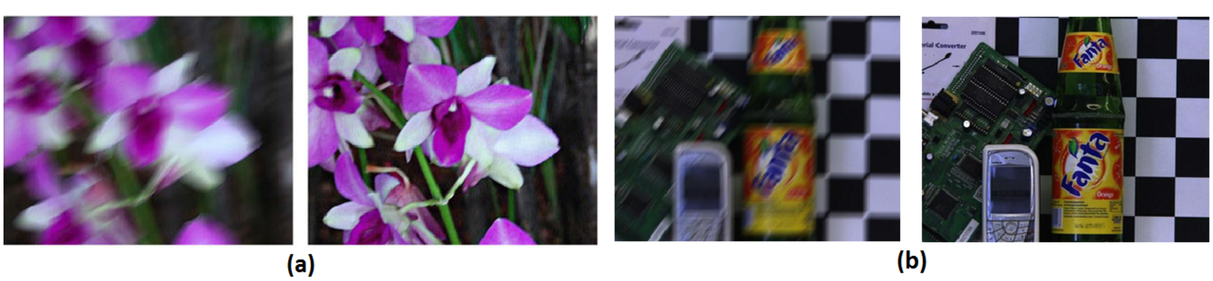


Figure 6.11 Deblurring on two challenging cases: the captured blurred images from [71] and corresponding IMD results.

6.6 Summary

In this work, a novel image restoration method has been proposed to remove camera motion blur from a single image by viewing deblurring as an optimisation process. The method does not involve estimation of the blur kernel or blur depth and achieves the deblurring iteratively. Our main contributions are an effective model for removing blur that accounts for its spatial distribution, and a local prior to suppress ringing artifacts. This model improves unblurred image estimation even with a very simple compensate function after a modified uni-variate optimization process is applied.

The proposed technique avoids the computation of blur depth parameter which is often erroneous. The successful results obtained with this technique is principally due to the optimization scheme that re-weights the relative membership values of neighboring pixels in current pixel value, over the course of the optimization. We have found that this re-weighting approach can work very accurately in case of horizontal uniform motion blur even if it is blurred by a large-size kernel.

The proposed technique was found to successfully deblur most motion blurred images. However, one failure mode occurs when the blurred image is affected by blur that is not shift-invariant, e.g., from slight camera rotation or non-uniform object motion. An interesting direction of future work is to explore the removal of non-shift-invariant blur using a general compensate function assumption.

Another interesting observation that arises from our work is that images, which are blurred with a very large-size kernel, contain more information than the original images. Our results show that for moderately blurred images, edge, color, and texture information can be satisfactorily recovered. A successful motion de-blurring method, thus, makes it possible to take advantage of information that is currently buried in blurred images, which may find applications in many imaging-related tasks, such as image understanding, 3D reconstruction, and video editing.

Chapter 7

CONCLUSIONS

In this work, we have proposed a local information based formulation to OF computation and relied on local statistics to obtain an accurate and robust flow. We have reported some investigations which showed that the performance of the formulation with NKD and KLD was found to be on par with the leading methods in terms of accuracy with an additional key strength of providing robustness to flow estimation which is lacking in current methods. Based on our experimental results, the following insights were observed for flow computation: NKD for good accuracy; NKD, KLD for robust flow computation and H, RD for fast computation. Thus, the proposed approach offers both robustness and accuracy without comprising either of them, albeit at a higher computational load. A fast computation of flow is feasible with H but with some trade off in accuracy. While the high computational cost may seem prohibitive for real time applications, the offsetting features are robustness to degradations and ability to extract structure from motion even from noisy data. This is of interest especially in applications such as echocardiography.

We demonstrated the performance of the proposed formulation on three different applications where one needs robust as well as accurate OF estimation. We achieved good results especially in motion edge detection as well as heart-beat rate estimation in ultrasonography. A new signal called the Temporal Flow Graph which describes the temporal behaviour of the landmark has been defined. HR estimation is shown to be simplified using the TFG. The proposed HR estimation can also be extended for in-vivo analysis for animal studies and for assessing fetal cardiovascular health. We also tried to automate the problem of color propagation in colorization of videos/films by using OF fields. The process of colorization remains a manually intensive and time consuming process. In this work, we have suggested a method that helps graphic artists to colorize films with less manual effort. We propose a framework which capitalises on the notion that not all frames will have maximum information together with the fact that frames of a scene are related by a motion field. Thus, an artist needs to color automatically selected most informative frames (1 per scene) which is subsequently propagated using OF fields. With the current framework, little more user effort is needed when the video contains more objects not all of which may be present in one frame such as capturing a scene with rotating camera or a still camera capturing a busy road (surveillance videos) scene. In such scenarios also, user effort for the proposed method is far less than that of other methods. Our future work aims at colorization of these kind of scenes with least user interactions.

We also proposed a method remove the effect of motion blur which is a common problem in capturing video with fast moving cameras. The proposed deblurring model improves unblurred image estimation even with a very simple compensate function after a modified uni-variate statistical information optimization process is applied. The proposed technique avoids the computation of blur depth parameter which is often erroneous. The successful results obtained with this technique is principally due to the optimization scheme that re-weights the relative membership values of neighboring pixels in current pixel value, over the course of the optimization. We have found that this re-weighting approach can work very accurately in case of horizontal uniform motion blur even if it is blurred by a large-size kernel. The proposed technique was found to successfully deblur most motion blurred images. However, one failure mode occurs when the blurred image is affected by blur that is not shift-invariant, e.g., from slight camera rotation or non-uniform object motion. An interesting direction of future work is to explore the removal of non-shift-invariant blur using a general compensate function assumption. Another interesting observation that arises from our work is that images, which are blurred with a very large-size kernel, contain more information than the original images. Our results show that for moderately blurred images, edge, color, and texture information can be satisfactorily recovered. A successful statistical information optimizing method, thus, makes it possible to take advantage of information that is currently buried in blurred images, which may find applications in many imaging-related tasks, such as image understanding, 3D reconstruction, and video editing.

Related Publications

1. V S Rao Veeravasarapu, Jayanthi Sivaswamy. *Local Information Flow*. Elsevier Jounal of Image and Vision Computing - 2012. (under review, submitted on 12, September 2012)

2. V S Rao Veeravasarapu and Jayanthi Sivaswamy. *Motion deblurring as optimisation*. In Proceedings of the Seventh Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP '10). ACM, 267-273.

3. V S Rao Veeravasarapu, Jayanthi Sivaswamy. *Fast and Fully Automated Film Colorization*. International conference on Signal Processing and Communications (SPCOM'12), 2012, IEEE.

Bibliography

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

[2] http://www.settheory.com/Glasspaper/motionedges.html.

[3] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, 1989.

[4] C. Auyeung and R.M.Mersereau. A dual approach to signal restoration. *In Information Sciences: Springer Series*, 23:21–56, 1991.

[5] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *IEEE 11th International Conference on Computer Vision.*, pages 1–8. IEEE, 2007.

[6] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International journal of computer vision*, 12(1):43–77, 1994.

[7] M. Ben-Bassat. f-entropies, probability of error, and feature selection. *Information and Control*, 39(3):227–242, 1978.

[8] M. Ben-Ezra and S. Nayar. Motion-based motion deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):689–699, 2004.

[9] A. Berlinet and L. Devroye. A comparison of kernel density estimates. *Publications de lInstitut de Statistique de lUniversité de Paris*, 38(3):3–59, 1994.

[10] M. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proceedings., Fourth International Conference on Computer Vision.*, pages 231–236. IEEE, 1993.

[11] S. Boltz, E. Wolsztynski, E. Debreuve, E. Thierry, M. Barlaud, and L. Pronzato. A minimum-entropy procedure for robust motion estimation. In *IEEE International Conference on Image Processing.*, pages 1249–1252. IEEE, 2007.

[12] T. Brox, C. Bregler, and J. Malik. Large displacement optical flow. 2009.

[13] T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE transactions on pattern analysis and machine intelligence*, pages 500–513, 2010.

[14] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnorr. Variational optical flow computation in real time. *IEEE Transactions on Image Processing*, 14(5):608–615, 2005.

[15] A. Bruhn, J. Weickert, and C. Schnorr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231, 2005.

[16] C. e. a. Bruser. Applying machine learning to detect individual heart beats in ballistocardiograms. In *EMBC-2010*, pages 1926–1929. IEEE, 2010.

[17] Z. Cernekova, C. Kotropoulos, and I. Pitas. Video shot segmentation using singular value decomposition. In *ICASSP'03*. IEEE, 2003.

[18] C. Chen. On information and distance measures, error bounds, and feature selection. *Information Sciences*, 10(2):159–173, 1976.

[19] C. Chubb and G. Sperling. Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A*, 5(11):1986–2007, 1988.

[20] D. Cunningham. Image motion deblurring. *Online Archive*, 2006.

[21] M. Donatelli, C. Estatico, A. Martinelli, and S. Serracapizzano. Improved image deblurring with antireflective boundary conditions and re-blurring. *Inverse Problems*, 22(6):2035–2053, 2006.

[22] T. Durrani and X. Zeng. Copula based divergence measures and their use in image registration. In *17th European Signal Processing Conference, Glasgow, UK*, pages 24–29, 2009.

[23] P. Estévez, M. Tesmer, C. Perez, and J. Zurada. Normalized mutual information feature selection. *IEEE Transactions on Neural Networks.*, 20(2):189–201, 2009.

[24] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. Freeman. Removing camera shake from a single photograph. *Acm Transactions On Graphics*, 25:787–794, 2006.

[25] H. Foroosh, J. Zerubia, and M. Berthod. Extension of phase correlation to subpixel registration. *Image Processing, IEEE Transactions on*, 11(3):188–200, 2002.

[26] A. Fraser and H. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33(2):1134–1140, 1986.

[27] B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, and N. Navab. Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming. In *Information Processing in Medical Imaging*, pages 408–420. Springer, 2007.

[28] D. Gokhale and S. Kullback. The information in contingency tables. 1978.

[29] E. Haber and J. Modersitzki. Intensity gradient based registration and fusion of multi-modal images. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 726–733, 2006.

[30] R. Haralick. The facet approach to optic flow. Technical report, DTIC Document, 1983.

[31] Y. He, A. Hamza, and H. Krim. A generalized divergence measure for robust image registration. *IEEE Transactions on Signal Processing.*, 51(5):1211–1220, 2003.

[32] A. Hero, B. Ma, O. Michel, and J. Gorman. Alpha-divergence for classification, indexing and retrieval. *Communication and Signal Processing Laboratory, Technical Report CSPL-328, U. Mich*, 2001.

[33] K. Hild, D. Erdogmus, and J. Príncipe. Blind source separation using renyi's mutual information. *IEEE Signal Processing Letters.*, 8(6):174–176, 2001.

[34] H. Ho and R. Goecke. Optical flow estimation using fourier mellin transform. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008.*, pages 1–8. IEEE, 2008.

[35] B. Horn and B. Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981.

[36] B. Horn and B. Schunck. " determining optical flow": A retrospective. *Artificial intelligence in perspective*, page 81, 1994.

[37] I. Jekova. Estimation of electrocardiogram peak frequency and periodicity during ventricular fibrillation. In *Computers in Cardiology, 2003*, pages 225–228. IEEE, 2003.

[38] J. Jia. Single image motion deblurring using transparency. *Proc. CVPR*, 2007.

[39] N. Joshi, T. Kadir, and S. Brady. Simplified computation for nonparametric windows method of probability density function estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (99):1–1, 2011.

[40] T. Kadota and L. Shepp. On the best finite set of linear observables for discriminating two gaussian signals. *IEEE Transactions on Information Theory.*, 13(2):278–284, 1967.

[41] T. Kailath. The divergence and bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology.*, 15(1):52–60, 1967.

[42] S. K. Kim and J. Paik. Out-of-focus blur estimation and restoration for digital auto-focusing system. *Electronics Letters*, 34(12):1217–1219, 1998.

[43] B. Kohler, C. Hennig, and R. Orglmeister. The principles of software qrs detection. *Engineering in Medicine and Biology Magazine, IEEE*, 21(1):42–57, 2002.

[44] S. Kullback. *Information theory and statistics*. Dover Pubns, 1997.

[45] A. Levin, R. Fergus, F. Durand, and B. Freeman. Image and depth from a conventional camera with a coded aperture. *Proc. Siggraph*, 2007.

[46] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 689–694. ACM, 2004.

[47] A. Likas and N. A. Galatsanos. Variational approach for bayesian blind image deconvolution. *IEEE Trans. om Signal Processing*, 52(8):2222–2233, 2004.

[48] B. Liu and A. Zaccarin. New fast algorithms for the estimation of block motion vectors. *IEEE Transactions on Circuits and Systems for Video Technology.*, 3(2):148–157, 2002.

[49] R. Liu and J. Jia. Reducing boundary artifacts in image deconvolution. *Proc. ICIP*, 2008.

[50] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International joint conference on artificial intelligence*, volume 3, pages 674–679. Citeseer, 1981.

[51] L. Lucy. Bayesian-based iterative method of image restoration. *Journal of Ast.*, 1974.

[52] B. Ma, A. Hero, J. Gorman, and O. Michel. Image registration with minimum spanning tree algorithm. In *Proceedings. International Conference on Image Processing.*, volume 1, pages 481–484. IEEE, 2000.

[53] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging.*, 16(2):187–198, 2002.

[54] S. Martin and T. Durrani. A new divergence measure for medical image registration. *IEEE Transactions on Image Processing*, 16(4):957–966, 2007.

[55] S. Martin, G. Morison, W. Nailon, and T. Durrani. Fast and accurate image registration using tsallis entropy and simultaneous perturbation stochastic approximation. *Electronics Letters*, 40(10):595–597, 2004.

[56] J. Martinerie, A. Albano, A. Mees, and P. Rapp. Mutual information, strange attractors, and the optimal estimation of dimension. *Physical Review A*, 45(10):7058–7064, 1992.

[57] B. McCane, K. Novins, D. Crannitch, and B. Galvin. On benchmarking optical flow. *Computer Vision and Image Understanding*, 84(1):126–143, 2001.

[58] J. Miskin and D. Mackay. Ensemble learning for blind image separation and deconvolution. *Advances In Independent Component Analysis*, pages 123–141, 2000.

[59] J. Modersitzki. *Numerical methods for image registration*. Oxford University Press, USA, 2004.

[60] M. E. Moghaddam and M. Jamzad. Finding point spread function of motion blur using radon transform and modelling the motion length. *Proc. ISSPIT*, pages 314–317, 2004.

[61] M. E. Moghaddam and M. Jamzad. Linear motion blur parameter estimation in noisy images using fuzzy sets and power spectrum. *EURASIP Journal on Advances in Signal Processing*, 2006.

[62] R. Neelamani, H. Choi, and R. G. Baraniuk. Fourier-wavelet regularized deconvolution for illconditioned systems. *IEEE Transactions On Signal Processing*, 52:418–433, 2004.

[63] D. noll. Restoration of degraded images with maximum entropy. *Journal of Global Optimization*, 10:91–103, 1997.

[64] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, pages 1226–1238, 2005.

[65] J. Pluim, J. Maintz, and M. Viergever. Image registration by maximization of combined mutual information and gradient information. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, pages 103–129. Springer, 2000.

[66] J. Pluim, J. Maintz, and M. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging.*, 22(8):986–1004, 2003.

[67] R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: Motion deblurring using fluttered shutter. *ACM Transactions on Graphics*, 25(3):795–804, 2006.

[68] A. Rav-Acha and S. Peleg. Two motion blurred images are better than one. *Pattern Recognition Letters*, 26:311–317, 2005.

[69] A. RRNYI. On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability-1961*, pages 547–561, 1961.

[70] G. Sapiro. Inpainting the colors. In *ICIP 2005*.

[71] S. Schuon and K. Diepold. Comparison of motion deblur algorithms and real world deployment. *IAC-2007*, B1:1–11, 2007.

[72] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. *Proc. SIGGRAPH*, 2009.

[73] Q. Shan, W. Xiong, and J. Jia. Rotational motion deblurring of a rigid object from a single image. *Proc. ICCV*, 2007.

[74] C. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[75] M. Sonka, V. Hlavac, R. Boyle, et al. *Image processing, analysis, and machine vision*, volume 2. PWS publishing Pacific Grove, CA, 1999.

[76] G. Sperling. Movement perception in computer-driven visual displays. *Behavior Research Methods & Instrumentation*, 1976.

[77] F. Steinbrucker, T. Pock, and D. Cremers. Large displacement optical flow computation withoutwarping. In *IEEE 12th International Conference on Computer Vision.*, pages 1609–1614. IEEE, 2009.

[78] C. Strecha, R. Fransens, and L. V. Gool. Multimodal and multiband image registration using mutual information. In *Proc. Theory and Applications of Knowledge-Driven Image Information Mining with Focus on Earth Observation (ESA-EUSC*, 2004.

[79] D. Sun, S. Roth, and M. Black. Secrets of optical flow estimation and their principles. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*, pages 2432–2439. IEEE, 2010.

[80] D. Sun, E. Sudderth, and M. Black. Layered Image Motion with Explicit Occlusions, Temporal Consistency, and Depth Ordering. 2010. NIPS.

[81] C. Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of statistical physics*, 52(1):479–487, 1988.

[82] P. Van Leeuwen and H. Kuemmell. Respiratory modulation of cardiac time intervals. *British heart journal*, 58(2):129, 1987.

[83] Vieira et al. Fully automatic coloring of grayscale images. *Image and vision computing*, 25(1):50–60, 2007.

[84] P. Viola and W. Wells III. Alignment by maximization of mutual information. In *Proceedings., Fifth International Conference on Computer Vision.*, pages 16–23. IEEE, 1998.

[85] X. Wang, P. Tino, M. Fardal, S. Raychaudhury, and A. Babul. Fast parzen window density estimator. In *International Joint Conference on Neural Networks (IJCNN).*, pages 3267–3274. IEEE, 2009.

[86] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. An improved algorithm for tv-l 1 optical flow. *Statistical and Geometrical Approaches to Visual Motion Analysis*, pages 23–45, 2009.

[87] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*, pages 1293–1300. IEEE, 2010.

[88] L. Yatziv and G. Sapiro. Fast image and video colorization using chrominance blending. *Image Processing, IEEE Transactions on*, 15(5):1120–1129, 2006.

[89] L. Yuan, J. Sun, L. Quan, and H. Y. Shum. Image deblurring with blurred/noisy image pairs. *Proc. Siggraph*, 2007.

[90] Z. Zhang. Estimating mutual information via kolmogorov distance. *IEEE Transactions on Information Theory.*, 53(9):3280–3282, 2007.