TEL AVIV UNIVERSITY

THE IBY AND ALADAR FLEISCHMAN FACULTY OF ENGINEERING

The Zandman-Slaner Graduate School of Engineering

Real Time Methods for Restoration, Enhancement and Analysis of Videos with Applications in Intelligent Transportation Systems

By

Barak Fishbain

THESIS SUBMITTED TO THE SENATE OF TEL-AVIV UNIVERSITY

in partial fulfillment of the requirements for the degree of

"DOCTOR OF PHILOSOPHY"

June 2008

THE IBY AND ALADAR FLEISCHMAN FACULTY OF ENGINEERING
The Zandman-Slaner Graduate School of Engineering

Real Time Methods for Restoration, Enhancement and Analysis of Videos with Applications in Intelligent Transportation Systems

By

Barak Fishbain

THESIS SUBMITTED TO THE SENATE OF TEL-AVIV UNIVERSITY
in partial fulfillment of the requirements for the degree of
"DOCTOR OF PHILOSOPHY"

Under The Supervision of Prof. Leonid P. Yaroslavsky – The Iby and Aladar Fleischman Faculty of Engineering, Tel-Aviv University
and
Assoc. Professor David Mahalel – The Faculty of Civil and Environmental Engineering, The Technion Haifa

June 2008

This work was carried out under the supervision

of

Professor Leonid P. Yaroslavsky

The Iby and Aladar Fleischman Faculty of Engineering, Tel-Aviv University

and

Assoc. Professor David Mahalel – The Faculty of Civil and Environmental Engineering,

The Technion Haifa

Acknowledgments

I would like to thank *Prof. Leonid Yaroslavsky* for his continuous support, scientific guidance, encouragement, and companionship that have guided me throughout my master and PhD degrees. I consider myself very lucky to have been guided by *Prof. Yaroslavsky* during my PhD research, and I hope that I have absorbed some of his modesty, integrity, and non-compromising quest for excellence. Additionally I would like to thank *Prof. David Mahalel*, the faculty of civil and environmental engineering, the Technion-Haifa, for his guidance, for being open for new ideas and disciplines and for opening, for me, new directions of research and new horizons.

I would like to thank *Dr. Zadok Hadas*, *Dr. Ishay Aloni* and *Dan Slasky*, El-Op Electro Systems, who helped me in the development of the turbulence compensation real-time algorithms. I would also like to thank my friends at the faculty for their friendship and help during this work. In particular I would like to thank *Shai Gepshtein* and *Alex Shtainman* who their works raised the initial ideas for the research; *Ofer Ben-Zvi*, *Alon Shtern* and *Ofer Shapira* for the help in coding and *Alex Agranovich* for his contribution with motion estimation techniques. *Ianir Ideses*, *Benny Salomon*, *Gil Shabat* and *Haggai Kirshner*, I thank you all for your friendship, and for all the help through endless discussions we had.

As typical for image processing research work, it required a lot of footage. I would like to thank the staff in Ayalon highway authority and Derech Eretz Operator, *Haggai Tal, Maya Elimelech* and *Royi Ninyo*, for their great help and support with this respect.

Special thanks are also in order to *Dr. Yaniv Harel*, *Dr. Doron Havazele*t and *Dr. Gabi Kedma*, the Israeli defense advanced research projects agency (MAFAT), and to *the Ministry of culture, sport and science*, *Or Yarok*, *ministry of transportation* and *the road safety authority* for their vote of confidence and financial support for this research.

I am most grateful to my loving family – My parents *Yossi* and *Livi* for their encouragement and constant support throughout my studies. I cannot imagine myself graduating without the love and support of my wife *Nili* who stood by me, encouraged me, and strengthen me, throughout difficult and sometimes frustrating parts of the research, and to my daughter *Yarden* who was born during my studies.

In memory of my brother Yarden

Abstract

Image and video enhancement is a core enabling technology in many fields. Nonetheless, several important questions such as multiframe restoration, enhancement and analysis of video streams, acquired in time-varying and unknown system and environmental conditions, remain unsolved. While many methods have been proposed throughout the years for solving multiframe restoration problems, well-established restoration methods exist for situations in which all sources of blur and degradation are known or easily predicted. When some of the parameters are unknown, however, the problem becomes much more difficult. The presented research addresses the following challenges: (i) efficient and robust motion estimation techniques; (ii) Real time methods for video stabilization and super-resolution in instable, due to camera and environmental noise, videos; and (iii) Motion-based scene analysis and reasoning.

At first motion estimation techniques are evaluated through a novel comparison framework. Based on the evaluation of these methods, an improvement, through numerical exact derivation, is suggested for the optical-flow class of motion estimation techniques.

Evaluation of the motion field and its statistical analysis allows a reliable segmentation of video frames into stable and moving components and subsequently stabilizing images, without harming real moving objects, and improving frames resolution. Along with the development of real-time methods for image stabilization and super-resolution, the potential and limitations of utilizing the motion field of instable sequences for super-resolution are sought. An important part in the process of resolution enhancement is signal reconstruction from sparse data accumulated from the set of randomly displaced image frames. The method used is improved by the theory of discrete signal reconstruction from sparse data.

Finally, based on the earlier stages, the accurate motion analysis along with image stabilization and resolution enhancement methods are utilized for providing means for reasoning the scene observed. This, for example, allows detection of irregularity of the motion in the scene. In traffic application, this corresponds to congestion or accidents.

The results presented are of both theoretical and practical interest and offer new efficient tools for substantial improvement of infrastructure of vision-based systems in general and of intelligent transportation systems in particular.

Contents

Table of Contents

List of Symbols and Abbreviations

Chapter 2.	
$\vec{\phi} = (\xi, \upsilon, \varsigma)$	Spatial-Temporal position of a pixel in a continuous coordinates
$\vec{p} = (x, y, t)$	Discredited spatial spatial-temporal position of a pixel
I	Image Intensity
(r, s)	Discrete Signal's transform domain spatial coordinates
$\vec{v} = (v_x, v_y)$	Velocity vector
$h(\vec{\phi}, \vec{u}; \theta(\tau))$	Time-varying system and/or environmental transfer function
$\theta(\tau)$	Set of time-varying parameters that determine the form of the transfer function
Chapter 3.	
$\dot{a}^{method}(x)$	Signal, *a*, derivative through *method* numerical differentiation scheme
$h^{method}(x)$	Digital convolutions kernel of *method* numerical differentiation scheme
$\eta^{method}(r)$	Point spread function of numerical differentiation *method*.
$e(\vec{p})$	Motion estimation error
Chapter 4.	
$DFM(\vec{p})$	Gray-level difference between running value of pixel $\vec{p}$ of incoming frame and its temporal median.
$MDM(\vec{p})$	Magnitude based likelihood that pixel $\vec{p}$ belong to an object in a real motion
$ADM(\vec{p})$	Angle distribution based likelihood that pixel $\vec{p}$ belong to an object in a real motion
$RMSM(\vec{p})$	Combined real-motion separation mask
a	Temporal window gray-level histogram
b	Temporal gray-level median value
c	Number of pixels in the temporal window, having the same intensity level as the median.
d	Number of pixels in the temporal window, having a lower intensity than the median.
Chapter 5.	
$C^{R,G,B}(x, y)$	R,G and B sampled output values of CCD cells covered by the Bayer filter
$Y^{R,G,B}(r, s)$	The Y component, in YCrCb color representation scheme, modulation transfer function for R, G or B pixels in the Beyer filter

Chapter 6.	
A_N	Vector of ***N*** samples of a continuous signal, acquired in a way that if all ***N*** samples are known, they are sufficient for representing the continuous signal.
$\widetilde{A}_K = \{a_{\tilde{k}}\}$	Vector of ***K*** signal's samples obtained in an irregular positions.
Φ_N	Matrix representation of an othronormal transform
$\Gamma_N = \{\gamma_r\}$	Signal's transform coefficients vector
$KofN_\Phi$	Matrix representation of signal's transform coefficients which are not zero for band-limited signals.
Chapter 7.	
q	Traffic flow
α	Average car size
$(x, y) \in \Omega$	Pixels located on a virtual line, Ω.
u_i	Individual vehicle speed
$\overline{u}_t$	Time mean speed
$\overline{u}_s$	Space mean speed
k	Traffic concentration

List of Figures

List of Tables

1. Introduction and Scientific Technological Background

Many methods have been proposed and studied for solving multiframe restoration problems. This chapter lays the concepts and principles of the research that have been carried till now and outlines the works done so far in the field. This survey was done in the light of applications for which multiframe restoration is utilized, without prior knowledge about the degradation characteristics.

1.1. Scope of the work

Recent development in video acquisition hardware has increased the number of applications which incorporate visual components. The development of such systems presents several challenges, among which one of the most challenging is visual enhancement. The challenges are getting bigger when the enhancement task requires scene analysis and the inference of a semantic description of the features extracted from the video stream (moving regions, trajectories, etc.). Even with the advancing camera and digital recording technology, there are many situations in which acquired images or videos suffer from severe visual degradations which hamper subsequent tasks such as scene reasoning and learning scene semantics.

The scope of the research described, is multiframe image restoration and enhancement concerned with the improvement of imagery acquired in the presence of time-varying degradations. The degradations can arise from a variety of factors - common examples include under-sampling of the image data, system aberrations and instabilities, and wave propagation through turbid media. Acquired images and videos are intended for visual analysis by a human operator. For this imagery are subjected to perfection and enhancement by means of digital processing. This is illustrated in Figure 1-1. The digital camera acquires the footage of the scene observed where the light propagates through a turbid instable medium. The output image is a product of a digital image processing of a set of images of the scene.

As depicted in Figure 1-2, the thesis is organized as follows. Chapter 3. establishes a novel comparison framework for motion estimation techniques and demonstrates that exact numerical derivation applied to optical flow yields superior results for estimating motion. Statistical analysis applied to this motion is used to

perform segmentation of the video frames into stable and moving components for image stabilization and improving frame resolution through super-resolution. This is elaborated on in chapters 4. and 5. respectively. Interpolation of data scattered due to randomly displaced image frames, completes the super-resolution process. To this end, the theory of discrete signal reconstruction from sparse data is formulated. Finally, the results of motion field analysis are applied to detect irregularity of scene motion with application to traffic. This is done through the evaluation, from traffic videos, of traffic flow theory parameters using the motion estimation, segmentation, stabilization and resolution enhancement methods developed within the scope of this thesis.

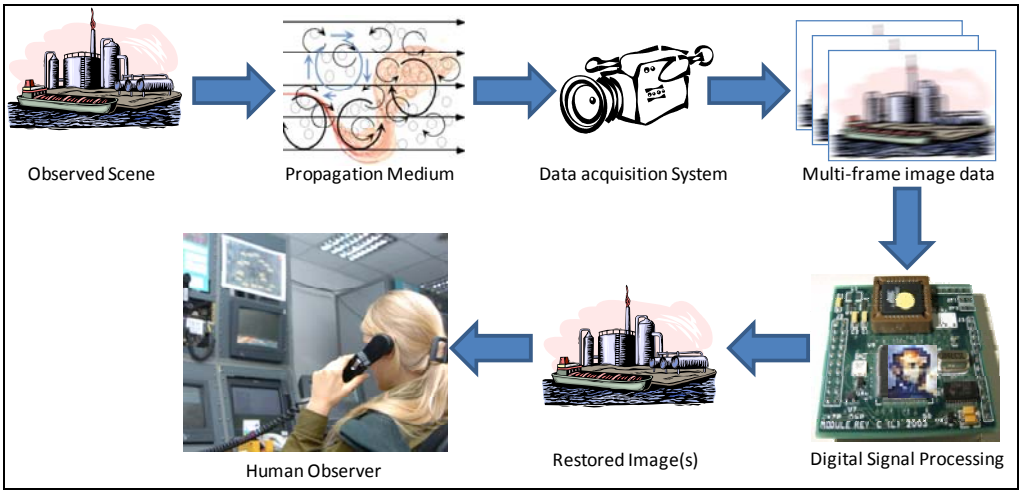


Figure 1-1 - A general scenario in which multiframe data are recorded and a restored image is produced through digital image processing.

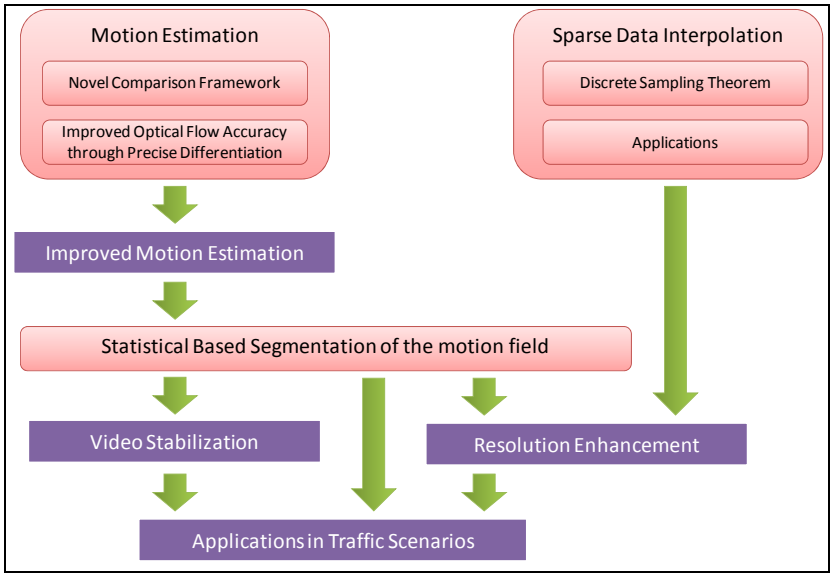


Figure 1-2 – Thesis road map

1.2. Image Acquisition through Turbid Media

Ideally, image quality in any acquisition systems would be limited only by the optical setup used. However many applications the major cause for image distortion is turbid propagation medium. Turbulence causes spatially and temporally chaotic fluctuations in the index of refraction of the propagation of which the light propagates through [1,2].

In long distance observation system and remote sensing applications, for example, light propagates long paths through the lower atmospheric regions (troposphere). The troposphere layer is in constant motion due to winds and local temperature variations [3]. These variations cause formation air pockets, which have a uniform index of refraction and can be modeled as spherically shaped turbulent cells in a range of sizes and densities (referred to as "turbulent eddies"). This causes small neighborhoods in the image sequences to chaotically move in different directions in different frames. Endoscopy [4] or underwater photography [5], where video is acquired through liquids, define another class of applications that imagery might be subjected to turbulence degradations. In all these cases, videos captured by optical sensors in the presence of turbulence are very unstable and degraded in their resolution and geometry. Visual analysis of such wavering output makes reliable detection and tracking of moving objects is practically impossible. This strongly motivates development of real-time methods for stabilization of turbulent videos.

Astronomical systems also suffer from atmospheric turbulence. In astronomical systems, the observed scene has a narrow field of view. Therefore the turbulence effect can be modeled by the convolution of the entire frame with a single, though random, point spread function (PSF). Adaptive optics methods were suggested for dealing with turbulence in astronomical imaging systems [6,7], where aberrations induced by the atmosphere were compensated using mechanical means. Classical adaptive optics, which uses a single deformable mirror, provides correction for a limited field of view (FOV).

Unlike astronomical systems, the images in the tasks described above, can be modeled by convolution of raw images with a space variant, pseudo-random PSF [8]. Therefore an adaptive optics solution for large FOV is desirable. Larger FOV corrections can be achieved by several deformable mirrors optically conjugated at various heights [9,10]. With that, multi-conjugate adaptive optics techniques require

complex structure and reconstruction processes, which make them unsuitable for operational systems.

Some methods for turbulence distortions compensation of images acquired through turbulence suggest finding the Modulation Transfer Function (MTF) of the observed scene and inverse filtering the output image to obtain turbulent free images [11,12,13]. Extracting the MTF from the video sequences requires some particular knowledge of the image content beforehand.

Naturally, atmospheric turbulence effects evolve with time. This evolution is characterized by a correlation time, τ_c. Qualitatively we can think of the atmosphere as generating a new field perturbation realization every τ_c second. If an imaging system exposes the image plane detector for a large number of atmospheric correlation times, the image is called a long exposure image. Short exposure imaging refers to the situation in which the exposure time is short enough to freeze the effects of the atmosphere, typically about a few milliseconds. Short exposure images were found to have a significantly different character than long exposure images. In particular, the short exposure images were broader than the point spread function of the optics alone, possessing high spatial frequency modulation not present in long exposure images. Further, the centroid of the intensity of the short exposure images was observed to move randomly about the image plane [14]. The problem of restoring images, degraded by motion blur, in long exposure systems, has been previously addressed in [15,16], where Hadar et al. formulated the mathematical expression of the degrading MTF. This work addresses scenarios where the camera is mechanically vibrating. Mechanical vibrations are typically characterized by swift and abrupt motion, with respect to the acquisition system's frame rate, hence long exposure conditions. Turbulent motion, on the other hand, is typically characterized by slow and moderating changes and therefore regarded as short exposure systems. In cases, where the acquisitioning conditions, do allow the system to be regarded as short exposure system, applying the inverse filter of the MTF described in [15,16], after applying the stabilization methods unfolded within the scope of this research, might improve the visual output. This is issue is also relevant when we consider resolution enhancement limitations, as described in Sect. 5.5.

Recently, several consumer and professional digital cameras incorporate anti-shake mechanisms. Those are using one of the following stabilization principles:

(a) Digital image correction

The camera motion is estimated from the digital input images captured by the camera and the movement correction is performed by digital processing of the camera images. The digital image correction methods can be classified into two groups: (i) full digital image correction, which is used in consumer video cameras [17] and (ii) video coding algorithms [18]. The first group deals only with compensation of certain types of translational image motion at rather low accuracy. The latter group uses more sophisticated algorithms, which can remove full first-order (affine) deformations between images in a sequence and assemble these aligned images within a single reference coordinate system to produce an image mosaic. The advanced algorithms applied, such as pyramid-based motion estimation and image warping [19] or fuzzy adaptive Kalman filtering [20], allow even subpixel accuracy, but require advanced processing hardware.

(b) Sensor based image correction

The camera motion is assessed with an external motion sensor and the movement correction is performed by digital processing of the camera images. The sensor-based image correction suffers from the fact, that the motion sensor is normally non-collocated with the actual image sensor. Therefore any mechanical distortions (misalignment, structural deformations, vibration) result in image motion measurement errors, which affect the final quality of the corrected image.

(c) Opto-mechatronic stabilization

The camera motion is compensated by a mechanically driven optical system. The opto-mechatronic stabilization requires a motion sensor and a mechanical actuator. This is a disadvantage in aspects of camera-size and power consumption. In addition, the system can not help for blurs caused by moving objects.

More recent digital image processing methods for turbulent video stabilization were suggested in [21,22]. The principal idea is to use, for reconstructing distortion compensated image frames, an adaptively controlled image re-sampling method based on the estimate of image local displacement vectors. Using those concepts, turbulence compensation algorithms which preserve genuine motion in the scene were suggested in [23,24,25]. However, those methods are computational complex and not suitable for real-time applications.

Most of known methods dealt with turbulence compensation by optical means. However, adaptive optics methods, as described earlier, are not adequate for long distance observation systems. Mentioned MTF methods, on the other hand, require

some prior knowledge about the scene that might not be available, and cannot cope with space variant distortions. The digital image processing techniques, presented, are not suitable for real-time applications.

1.3. Resolution Enhancement and Super-resolution

Super-Resolution (SR) is a general term for a set of methods for increasing image (or video) resolution. Most of SR techniques are based on the fact that no real observation platform can be absolutely stationary; hence consequent frames can differ also due to small local or global instabilities in the image plane caused by unsteadiness, such as atmospheric turbulence, in light propagation media. At the same time, the video frames usually contain information on stable scenes, which is not supposed to change from frame to frame. Therefore observed image instabilities can be, if properly evaluated, used for enhancing image resolution beyond the limits defined by the acquisition camera sampling ratio.

Generally, SR techniques can be divided into three main stages:

- Determination with sub-pixel accuracy the location of every pixel in each low-resolution (LR) frames; this can be done by applying an optical flow or a motion estimation algorithm to the input LR video.
- Accumulation of the data to a new image to obtain an up-sampled image.
- Interpolation of the new image for obtaining a complete image and to remove aliasing.

SR principles and general multi-channel image recovery are detailed in the works of T. J. Schultz [1], S. Srinivasan and R. Chellappa [26], Galatsanos, Wernick and Katsaggelos [27] and R.R. Schultz [28]. Several researchers treat the problem of high resolution image recovery by designing an efficient multiframe filtering algorithms, that account for both intra-frame (spatial) and inter-frame (temporal) correlations, for restoring image sequences that are degraded both by blur and noise [29,30]. Others have formulated solutions to global motion problems, usually from an application perspective [31,32,33,34,35,36,37,38,39,40,41,42]. These can be broadly classified as feature-based and flow-based techniques.

Feature-based methods extract and match discrete features between frames, and the trajectories of these features are fit to a global motion model. In flow-based

algorithms, the optical flow of the image sequence is an intermediate quantity that is used for determining the global motion.

Irani and Peleg [33], suggested a SR algorithm based on registration of several images and projecting their shifted samples into a new "initial guess" to create a SR image. SR for infrared image using global motion is presented in [34], where the forward looking infrared (FLIR) array detector is located on a moving object such as a car or an aircraft. In the latter image registration is used to place all the samples on a new grid and produce super resolved image. Other SR methods who rely on global motion (non-stationary camera) and image registration can be found in [39,41,42].

A different kind of SR obtained from multiple images, this time through zooming instead of motion was suggested by Li [43]. In his work, the input LR frames are images of the same object zoomed. In each frame, the samples are slightly taken in different position. Fusing all those samples to a new denser grid creates a "synthetic zoom" image. Resembling idea was introduced by Nayar et al. who proposed the "jitter camera" for achieving super-resolution [44]. This video camera that samples the space-time volume at different locations without introducing motion blur. This is done by instantaneously shifting the detector (e.g., CCD) between temporal integration periods, rather than continuously moving the entire video camera during the integration.

The impact of image wrapping methods on the quality of image resolution enhancement was studied by Chiang and Boult [45]. Image warping requires the underlying image to be "resampled" at non-integer locations; hence, it requires image reconstruction. When the goal of warping is to produce output for human viewing, only moderately accurate image intensities are needed. In these cases, techniques using bi-linear interpolation have been found sufficient. However, as a step for applications such as super-resolution, the precision of the warped intensity values is often important. For these problems, bi-linear image reconstruction may not be sufficient. To this end, a warping approach, based on the integrating resampler [46] which warps the image while both enforcing the underlying image reconstruction and satisfying the imaging consistent constraint [47], was suggested. The imaging consistent constraint requires that the image reconstruction yields a function which, when convolved with the imaging system's point-spread function (PSF), is consistent with the input image.

In applications where the acquiring is held under turbulent conditions, the image sampling grid defined by the video camera sensor can be considered to be chaotically moving over a stationary image scene. As described in 1.2, in turbulence-degraded videos, consequent frames of a stable scene differ only due to atmospheric turbulence-induced local displacements between images. This phenomenon allows generating images of such scenes with larger number of samples than that provided by the camera if consecutive image frames are combined by means of their appropriate re-sampling. Resolution enhancement from turbulent videos was investigated by Charnotskii [48] who showed preliminary feasibility for SR in physical experiments taken under laboratory conditions. These experiments, based on taking photos of an object through turbulent medium, showed a certain potential for SR by using multiple turbulence distorted frames. Fraser *el al.* [49, 50] have suggested similar method for increasing image resolution by using turbulence distorted videos. In their work, the SR image is obtained by first creating relatively blurred reference frame of temporal average over the LR distorted frames, and then calculating the local shifts between each LR frame and the reference frame using local correlation or optical flow. Then, the reference frame is re-calculated using the corrected frames, and the process is repeated iteratively. This suggested algorithm requires high system resources and is not suitable for real time applications. Also, no interpolation method is used to remove all the aliasing and convergence is not discussed.

1.4. Sparse Data Interpolation

An important part of the process of resolution enhancement is signal reconstruction from sparse data accumulated from the set of randomly displaced image frames. The theory of discrete signal reconstruction from sparse data is addressed in a number of publications. In this chapter, the discrete sampling theorem is formulated.

Images are usually represented in computers in a form of samples on a uniform rectangular sampling grid. However, in many applications, sampled data are collected in irregular fashion and/or it may frequently happen that some samples of a regular sampling grid are lost or unavailable. In these cases, it is required to convert irregularly sampled signals to regularly sampled ones or restore missing data. Typical examples are filtering “salt &pepper”-type noise in images transmitted through communication channels with error detection coding; reconstruction of surface

profiles from sparse data; wave front profile reconstruction from interferometric data in geophysics and optical metrology; restoration of image sequences acquired in the presence of camera or object vibrations or through a turbulent medium, and image super-resolution from multiple chaotically sampled frames, to name a few.

There are two approaches to treat this problem. One approach is empirical in nature and is based on simple numerical interpolation procedures. For example, in Shepard's interpolation method, which is one of the first methods for interpolating a 2D irregularly sampled signal, a missing image sample is recovered by a weighted interpolation of close known samples [51]. The weights are inversely proportional to the distance of the samples from the missing one. Hence, closer samples will have larger weights and will be more influential in determining the missing samples. A review of these methods can be found in [52].

The second approach is based on generalizations of the classical uniform sampling. It is assumed that the given samples are obtained from a continuous signal that belongs to a certain approximation subspace M (e.g., band-limited signals, splines subspaces, etc.) of the parent Hilbert space (usually, L^2 Hilbert space of finite energy functions). An interpolation procedure has to determine a continuous signal that satisfies two constraints: 1) the signal has to be in the subspace M and 2) its values are equal to the given samples in the same positions. Conditions for existence and uniqueness are dependent on the signal model (underlying approximation subspace) and the set of given samples. For the band-limited case, Landau proved that a necessary and sufficient condition for the unique reconstruction of a continuous band-limited 1D signal with bandwidth W from its irregularly spaced samples is that the density of its samples should exceed the Nyquist rate $1/W$ [53]. It is also shown that this condition is necessary for D-dimensional signals with band limited Fourier spectrum. These results have been generalized to other shift-invariant subspaces by Aldourbi and Grochenig [54]. A comprehensive presentation of this approach can be found in [55]. Continuous band-limited signal reconstruction requires, however, an infinite number of samples. In addition, the sinc function used in the interpolation process is slowly decaying. Splines subspaces have proved to be an attractive alternative approximation model [56]. However, due to their localized nature, their use is limited for the recovery of large gaps in data. A practical numerical algorithm for interpolation and approximation of 2D signals based on multilevel B-Splines is given by Wolberg *et al.* [57]. The algorithm approximates 2D functions from sparse

data by an iterative procedure based on lattice control points. At each iteration, the values of available samples are preserved (if possible) or approximated. At the next iteration, a denser grid of control points is created to approximate the reconstruction error, and the process continues iteratively. Similar Spline-based algorithm was suggested by Margolis and Eldar [58] which uses, for interpolation, non-uniform Splines.

Hasan and Marvasti have suggested a method for recovery of missing samples, in the DCT transform domain [55]. In Ref. [59], the problem of nonuniform sampling in Fourier domain in multidimensional polar coordinates is addressed in connection with image reconstruction from projections. In yet another publication [60] Averbuch and Zheludev discuss image reconstruction from projections with omissions using biorthogonal wavelets over-complete bases functions.

While a significant amount of research was put on this class of problems, each of these methods discussed the approximation of discrete signals, specified by their sparse samples, with respect to one or two signal representation basis functions. Hence, no generalize framework has been formulated to describe all methods by coherent formulation.

1.5. Intelligent Transportation Video Systems Applications

Visual surveillance for traffic systems requires short processing time, low processing cost and high reliability [61]. Under those requirements, image processing technologies offer a variety of systems and methods for Intelligence Transportation Systems (ITS) as a platform for traffic Automatic Incident Detection (AID). An extensive survey of the methods can be found in [62]. According to different traffic data sources, there exist two classes of AID methods mainly studied: one is based on inductive loops, radars, infrared sonar and microwave detectors [63] and the other is based on video images [64]. The first class of methods suffers from drawbacks in that they are expensive to install and maintain and they are unable to detect slow or stationary vehicles. Video sensors, on the other hand, offer a relatively low installation cost with little traffic disruption during maintenance. Furthermore, they provide wide area monitoring allowing analysis of traffic flows and turning movements, speed measurement, multiple-point vehicle counts, vehicle classification and highway state assessment, based on precise scene motion analysis.

The vast majority of traffic videos analysis methods segment the motion field into objects' trajectories. Hence, scene's motion reasoning is achieved by aggregating the specific objects' motion information [65,66,67,68]. The different trajectories are then analyzed to detect slow motion, swift changes in speed or trajectories interference for generating alerts.

Examination of each individual object's motion in the scene is a complicated task, which might be not practicable under real-time constrains. When the acquired scene is a complex enough, as, for example, highway interchange or junctions, those problems become a real obstacle on the way to real-life implementations. Some of the methods and real life applications try to overcome this problem by limiting the examination area in the image to a single line or small neighborhood defined by the user [66]. The shortcoming of those solutions is that events outside those areas will not be detected. Additionally those methods are fundamentally limited by the accuracy that can be achieved by the motion estimation method used. The latter problem can be dealt with by regarding the entire motion field rather than the different vehicles' trajectories. A recent work has suggested analyzing the entire motion field by wavelet decomposition and not by motion segmentation [69]. However the rigid structure of the wavelet basis functions might not be appropriate for the task in hand.

Traffic models, which describe the relationship among traffic stream characteristics, such as flow, speed and concentration, are the foundations of traffic research for the last 40 and 50 years [70]. Those models consist of mathematical micro- and macroscopic description of road conditions. While a consolidate traffic theory is available, none of the image processing techniques, presented for traffic analysis problems, has suggested utilizing traffic flow models into traffic video systems.

1.6. Outline of the Thesis

The presented research focuses on developing restoration, enhancement and analysis methods for processing, in real-time, video streams, acquired in time-varying and unknown system and environmental conditions, in order to allow better presentation and characterization of the actions taking place and, in particular, to infer whether they present a threat that should be signaled. While many methods have been proposed throughout the years for solving multiframe restoration problems, well-established restoration methods exist for situations in which all sources of blur and

degradation are known or easily predicted [1,29,71,72,73,74,75,76]. When some of the parameters are unknown, however, the problem becomes much more difficult because, in addition to the object intensity, the unknown system parameters must also be estimated. This problem is also referred to as the blind multiframe restoration problem. The mathematical foundations of the multi-frame restoration problem are given in Chapter 2.

Efficient real motion extraction is mandatory for successful restoration and resolution enhancement and reliable scene reasoning. Numerous motion estimation techniques have been suggested for this goal in the last 20 years [77,78,79,80,81,82,83,84,85,86,87,88,89] to name a few. One of the goals of this research is the evaluation of these techniques and, based on this to develop one robust enough to atmospheric and optical noise and noise caused by vehicle or camera movements. Nowadays, motion estimation methods' accuracy is tested on very few "ad hoc" test sequences [90] with no quantitative evaluation framework [91]. Those commonly used sequences do not necessarily represent typical scenarios. Additionally, those methods are not oriented to work under real-time constrains. This is elaborated, in Chapter 3.

One common type of applications where multi-frame restoration is utilized is remote sensing (see Sect. 1.2). In the presented research, digital image processing techniques are further elaborated and improved upon in order to obtain turbulence compensation under real-time constrains. This is thoroughly described in Chapter 4.

As described in Sect. 1.3, turbulent degraded video sequences are suitable candidates for SR applications. The idea of making a profit from atmospheric turbulence-induced image geometrical spatial/temporal degradations to compensate image sampling artifacts and for static and dynamic SR applications (for the definition of static and dynamic SR see [92]), is elaborated in Chapter 5. Additionally, Chapter 5. addresses the fundamental question "What can be achieved?" or "What are *the potentials and limitations* of SR under different constraints, such as number of input frames, camera's PSF, etc. [93].

Chapter 6. addresses signal reconstruction from irregular samples and elaborates a new approach to the problem of optimal recovery of signals from irregularly sampled or sparse data.

Finally, the motion fields and the stabilization and resolution enhancement algorithms are evaluated to provide means for reasoning the scene observed for

detection of irregularity of the motion in the scene. The utilization of accurate motion analysis along with classical traffic theory in video-based traffic systems allows both improving system performance as well as giving tools for traffic researchers. Chapter 7. presents traffic applications for the technology developed within the scope of this research and the great potential of integration of classical traffic flow models into video-based intelligent transportation systems.

The results have been achieved in the above line of work are of both theoretical and practical interest since they contribute to further development of advanced video processing methods and, specifically, offer new efficient tools for substantial improvement of infrastructure of vision-based intelligent traffic systems for traffic monitoring and alerting on traffic anomalies and law enforcement. The discussion can be found in Chapter 8. Conclusions are formulated in Chapter 9.

2. Mathematical Formulation

This chapter lays the mathematical formulations of the multiframe restoration problem. First it defines the environment time-varying and system's sampling modulation transfer functions. Then, the mathematical depiction of the restoration problem for those systems is formulated. This formulation shows that under reasonable set of assumptions, the turbulence compensation problem and the resolution enhancement problem can be described through the same mathematical formulation. This important conclusion is exploited in subsequent chapters.

2.1. Notations and Preliminaries

Both continuous and discrete representations of motion and images will be used. Let $\vec{\phi} = (\xi, \upsilon, \tau)$ be a spatial-temporal position of a pixel in a continuous coordinates, i.e. $\vec{\phi} \in R^3$ within image limits, and let I denote image intensity. Image intensity, $I(\vec{\phi}) \in R$, is referred to as a nonnegative function that represents an object's ability to reflect or emit light (or other electromagnetic radiation) and is limited by the dynamic range of the sensing device (e.g., CCD). The continuous signal's transform domain spatial coordinates are given by (f_ξ, f_υ).

Before images can be manipulated digitally, they have to be sampled and quantized. Let $\vec{p} = (x, y, t)$ be a discreted spatial-temporal position in the image that corresponds to $\vec{\phi}$. The triplet (x, y, t) belongs to a 3-D sampling grid, for example a 3-D lattice. The same symbol *I* is used for continuous and quantized intensities; the nature of *I* can be inferred from its arguments. The discrete signal's transform domain spatial coordinates are given by *r* and *s*.

Motion can be described by a velocity vector $\vec{v} = (v_x, v_y)$, whereas $\vec{v}(\vec{p})$ is a velocity at position $\vec{p}$ and $\vec{v}_t$ denotes a velocity or motion field, hence the set of all velocity vectors within the image, at frame *t*.

2.2. Acquisition Systems' Transfer Function

The imaging problems discussed within the scope of the research, all involve the detection and processing of electromagnetic fields after reflection or emission from a

remote object or scene. Furthermore, the applications considered here are all examples of planar incoherent imaging, wherein the object or scene is characterized by its incoherent reflectance or emission function I . The central task of a multiframe image restoration problem, then, is the estimation of this intensity function from a sequence of noisy, blurred images. The need for image restoration is, in general, motivated by two factors: (i) detector sampling; and (ii) system and environmental deformations. These two stages of image formation are described in the proceedings sections.

2.2.1. System and Environmental Transfer Function

In all imaging applications, the signal available for detection is not the image intensity *I*. Instead, *I* is blurred by the imaging system, and the observed signal is:

$$\bar{I}\left(\vec{\phi},\tau\right)=\int h_e\left(\vec{\phi},\vec{u};\theta(\tau)\right)I(\vec{u})d\vec{u} \tag{2-1}$$

where $h_e\left(\vec{\phi},\vec{u};\theta(\tau)\right)$denotes the time-varying system and environmental PSF; $\vec{u}=(u_1,u_2)$is a two-dimensional spatial variable in the aperture plane ; $\bar{I}\left(\vec{\phi}\right)$denotes the time-varying resulted continuous-domain intensity; and $\theta(\tau)$denotes a set of time-varying parameters that determine the form of the PSF. In optical systems $h_e\left(\vec{\phi},\vec{u};\theta(\tau)\right)$ is affected by the system's diffraction, optical aberrations and the heterogeneity of the propagation medium.

Wave propagation through an inhomogeneous medium, such as turbid media, induces image distortions, due to the medium's refractive index. When taking the system's diffraction and optical aberrations into consideration, turbid medium's transfer function can be modeled in the following manner [2]:

$$h(\vec{\phi})=\left|\int A(\vec{u})e^{-j\frac{2\pi}{\lambda f}\left(\vec{u}\cdot\vec{\phi}\right)}e^{j\theta(\vec{u},\tau=\tau_0)}d\vec{u}\right|^2 \tag{2-2}$$

where $A(\vec{u})$is the system's aperture function, λ is the nominal wavelength of the detected radiation, and *f* is the system focal length [94]. The notation $\vec{u}\cdot\vec{\phi}$ denotes the inner product operation, and it is defined for two-dimensional variables as:

$$\vec{u}\cdot\vec{\phi}=u_1\xi+u_2\upsilon \tag{2-3}$$

Finally, $\theta(\vec{u})$ is the aberration function which can vary with time. We will model the system and environmental PSF as the (possibly) space-variant function $h(\vec{\phi})$, and note that this model captures diffraction, system aberrations, time-varying translations and rotations, and environmental distortions such as turbulence. The parameter $\theta(\vec{u})$, may be a simple vector parameter, or a more complicated parameterization of a three-dimensional function. Very frequently $\theta(\vec{u})$, will not be well known or predicted, and the identification of this parameter can be one of the most challenging aspects of a multiframe image restoration problem.

2.2.2. Sampling Transfer Function

The detection of imagery with discrete detector arrays results in the measurement of the (time-varying) sampled intensity:

$$\bar{I}(\vec{p}) = \int I(\vec{\phi}) h_s(\vec{\phi}, \vec{p}) d\vec{\phi} \tag{2-4}$$

where $I(\vec{\phi})$is the observed object, $h_s(\vec{\phi}; \vec{p})$ is the sampling device's transfer function of each point in the image and $\bar{I}(\vec{p})$is the discrete-domain intensity that results from sampling of the continuous domain.

The response function, $h_s(\vec{\phi}; \vec{p})$, for most digital cameras is often of the form:

$$h_s(\vec{\phi}, \vec{p}) = \begin{cases} 1, & \vec{\phi} \in \Omega_{\vec{p}} \\ 0, & \vec{\phi} \notin \Omega_{\vec{p}} \end{cases} \tag{2-5}$$

where $\Omega_{\vec{p}}$, denotes the spatial region of integration for the detector element located at $\vec{p}$. The regions of integration for most detectors are typically square or rectangular regions centered about the detector locations.

2.2.3. The Combined Modulation Transfer Function

The combined effects of blur and sampling are modeled as:

$$\begin{aligned} \bar{I}(\vec{p}) &= \int h_s(\vec{\phi}, \vec{p}) \int h_e(\vec{\phi}, \vec{u}; \theta(\tau)) I(\vec{u}) d\vec{u} d\vec{\phi} = \\ &= \int \left[\int h_s(\vec{\phi}, \vec{p}) h_e(\vec{\phi}, \vec{u}; \theta(\tau)) d\vec{\phi} \right] I(\vec{u}) d\vec{u} = \\ &= \int h(\vec{p}, \vec{u}; \theta(\tau)) I(\vec{u}) d\vec{u} \end{aligned} \tag{2-6}$$

where $h(\vec{p},\vec{u};\theta(\tau))$ denotes the mixed-domain (continuous-discrete) PSF and is given by:

$$h(\vec{p},\vec{u};\theta(\tau)) = \int h_s(\vec{\phi},\vec{p})h_e(\vec{\phi},\vec{u};\theta(\tau))d\vec{\phi} \tag{2-7}$$

2.3. The Restoration Problem

Stated simply, the restoration problem is one of estimating the image intensity $I(\vec{p})$ from the multiframe data set $\{I(\vec{p}); \forall \vec{p} = (x, y, t = 1,2,3,4,..., K)\}$. Statistical inference problems such as those encountered in multiframe image restoration are frequently classified as ill-posed problems [95]. An image restoration problem is ill posed, in the classical sense of Hadamard, if it does not have a unique solution. Multiframe image restoration problems that are formulated on infinite-dimensional parameter spaces are almost always ill posed, and their ill-posed nature is usually due to the discontinuity of the solution. Problems that are formulated on finite-dimensional spaces (as ours is here) are frequently well posed. However, these problems are usually ill conditioned or badly behaved and are frequently classified as ill posed even though they are technically well posed.

For problems that are ill posed or practically ill posed, the original problem's solution is often replaced by the solution to a well-posed (or well-behaved) problem. This process is referred to as regularization, and the basic idea is to change the problem in a manner such that the solution is still meaningful but no longer badly behaved [96]. The consequence for multiframe restoration problems is that we do not seek to match the measured data perfectly. Instead, we settle for a more stable - but inherently biased - image estimate.

In situations where the parameters $\theta(\tau)$ that characterized the MTF functions are time-varying and unknown or not easily predicted, the unknown parameters must be eliminated or estimated. Previously suggested methods for solving such multiframe blind restoration problems, have been tailored for specific blur models. In many cases it is assumed that the data collection interval for each frame is short compared with the fluctuation time of the parameter τ, so that each frame is recorded with temporal-wise constant transfer function, $h(\vec{p},\vec{u};\theta(\tau=\tau_0))$ as in (2-2). This suggests that image degradations due to the turbulence effect manifest themselves in image geometrical local deformations, rather than motion blur. In the light of this, (2-1) can be written in the following format:

$$\bar{I}\left(\vec{\phi}\right)=\int h\left(\vec{\phi}-A^{-1}\left(R\left(\tau=\tau_0\right)\right)\cdot\left(\vec{u}+\Delta\left(\tau=\tau_0\right)\right)\right)I\left(\vec{u}\right)d\vec{u} \tag{2-8}$$

where $\Delta\left(\tau=\tau_0\right)$ represents a two-dimensional translation on the aperture plane at time τ_0, and

$$R\left(\tau\right)=\begin{bmatrix}\cos\left(\alpha_\tau\right) & -\sin\left(\alpha_\tau\right)\\ \sin\left(\alpha_\tau\right) & \cos\left(\alpha_\tau\right)\end{bmatrix} \tag{2-9}$$

is a time-varying rotation matrix (at angle α_τ), representing the translations and rotations between the sensor and the scene at time τ_0.

So the shift variant PSF can be written as:

$$h\left(\vec{\phi}\right)=h\left(\vec{\phi}-A^{-1}\left(R\left(\tau=\tau_0\right)\right)\cdot\left(\vec{u}+\Delta\left(\tau=\tau_0\right)\right)\right) \tag{2-10}$$

and the parameters characterizing the PSF are then $\theta\left(\tau\right)=\left(\Delta\left(\tau\right),\alpha\left(\tau\right)\right)$.

While equation (2-10) was derived for turbulent scenarios, it resembles the PSF of micro-scanning applications [1]. Micro-scanning systems introduce small controlled global movements of the scene observed on the aperture plane. Those systems are designed to produce images with higher spatial resolution than that defined by the camera sampling grid. Following this notion, one can derive that the core base of the SR techniques designed for micro-scanning applications can be utilized, with the proper adjustments, in systems subjected to turbulence noise.

3. Motion Estimation in Videos

Efficient real motion extraction is mandatory for the success of the subsequent task: stabilization, resolution enhancement and video analysis. This chapter describes the fundamentals of motion estimation techniques and evaluates these techniques. Nowadays, motion estimation methods' accuracy is tested on very few "ad hoc" test sequences. In this chapter a novel, more general, comparison framework is presented. Finally, an innovative approach for achieving improved optical flow accuracy through precise differentiation is suggested.

3.1. Motion Estimation Techniques

A video sequence is a much richer source of visual information than a still image. This is primarily due to the capture of motion; while a single image provides a snapshot of a scene, a sequence of images registers the dynamics in it. Motion is important for video processing and compression for two reasons. First, motion carries a lot of information about spatiotemporal relationships between image objects. This information can be used in such applications as traffic monitoring or security surveillance, for example to identify objects entering or leaving the scene or objects that just moved. Secondly, image properties, such as intensity or color, have a very high correlation in the direction of motion, i.e., they do not change significantly when tracked in the image (the color of a car does not change as the car moves across the image). This can be used for the removal of temporal video redundancy; in an ideal situation only the first image and the subsequent motion have to be transmitted. It can be also used for general temporal filtering of video. In this case, one-dimensional temporal filtering along a motion trajectory, e.g., for noise reduction or temporal interpolation, does not affect the spatial detail in the image.

The above applications require that image points be measured how they move, this task is commonly referred to as motion estimation. Note that only two-dimensional (2-D) motion of intensity patterns in the image plane, often referred to as apparent motion, are considered. Three-dimensional (3-D) motion of objects, on a 3D model space, is not treated.

Methods explicitly reducing the number of bits needed to represent a video sequence will be referred to as video compression techniques. For example, motion compensated hybrid (predictive/DCT) coding is exploited today in all video compression standards, i.e., H.261, H.263, MPEG-1, MPEG-2, MPEG-4. In contrast, methods that do not attempt such a reduction but transform the video sequence, e.g., to improve quality, will be considered to belong to video processing methods. The discussion of motion in this chapter will be carried out from video processing point of view.

The above classification is important from the point of view of the goals of motion estimation, which in turn influences the choice of models and estimation criteria. In the case of video compression, the estimated motion parameters should lead to the highest compression ratio possible (for a given video quality). Therefore, the computed motion need not necessarily resemble the *true* motion of image points as long as some minimum bit rate is achieved. In video processing, however, it is the *true* motion of image points that is sought. Two families of motion estimation methods are available: block matching [77-83] and optical flow [84-89]. The principals of those techniques are detailed in Sect. 3.2 and Sect. 3.3, correspondingly.

3.2. Block Matching Methods

The underlying supposition behind motion estimation is that the patterns corresponding to objects and background in a frame of video sequence move within the frame to form corresponding objects on the subsequent frame. The idea behind block matching is to divide the current frame into a matrix of 'macro blocks' that are then compared with corresponding block and its adjacent neighbors in the previous frame to create a vector that stipulates the movement of a macro block from one location to another in the previous frame. This movement calculated for all the macro blocks comprising a frame, constitutes the motion estimated in the current frame. The search area for a good macro block match is constrained up to a certain number of ***n*** pixels on all fours sides of the corresponding macro block in previous frame. This ***n*** is called as the search parameter. Larger motions require a larger ***n***, and the larger the search parameter the more computationally expensive the process of motion estimation becomes. Usually the macro block is taken as a square of side 16 pixels, and the search parameter ***n*** is 7 pixels. The idea is represented in Figure 3-1. The matching of one macro block with another is based on the output of a cost function.

The macro block that results in the least cost is the one that matches the closest to current block. There are various cost functions, of which the most popular and less computationally expensive is Sum of Absolute Difference (SAD) given by equation (3-1). Another cost function is Mean Squared Error (MSE) given by equation (3-2).

$$SAD = \sum_{\vec{p}\in\Omega} \left| I(\vec{p}) - I(\vec{p}') \right| \tag{3-1}$$

$$MSE = \frac{1}{N^2} \sum_{\vec{p}\in\Omega} \left(I(\vec{p}) - I(\vec{p}') \right)^2 \tag{3-2}$$

where N is the size of the macro block and a $I(\vec{p})$ and $I(\vec{p}')$ is the intensity levels in a corresponding neighborhoods Ω in two sequential frames.

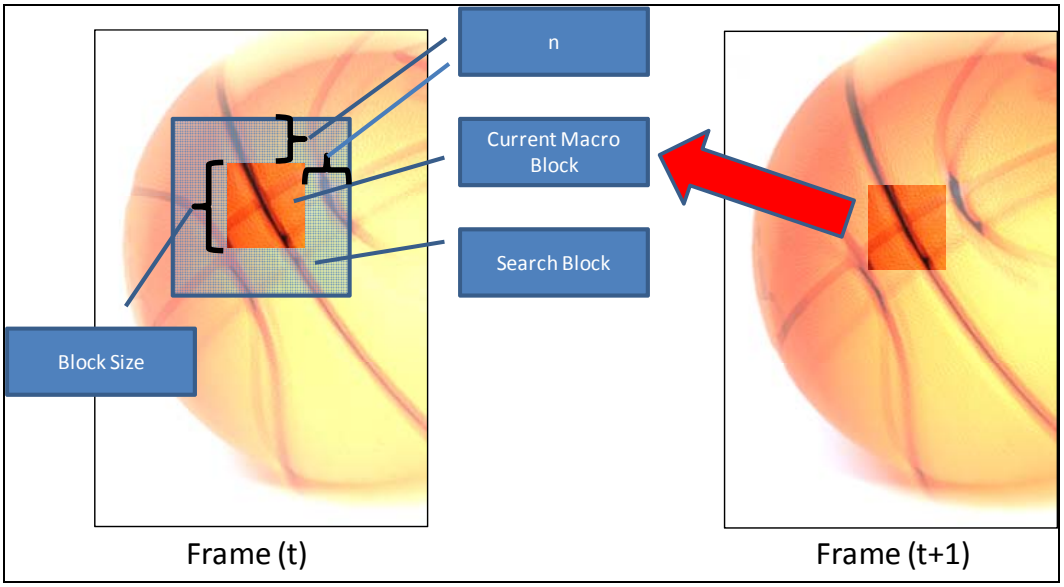


Figure 3-1 – Block matching – a macro block is extracted from a given frame and is searched within a search block in the previous frame. The parameter n determines the size of the search area.

Recently several new block-matching high efficiency algorithms were presented [80-83]. Those methods, by applying certain assumption on of the error function, such as smoothness and global minima, reduce the computational complexity. The number of checking points, examined within the search area, is reduced using efficient checking patterns, such as diamond or spiral. Additionally these techniques suggested several highly likely predictors. Finally, by introducing very reliable early-stopping criteria, the search is terminated before all sampling points' error needs to be computed. However, those improved techniques, as described in Sect. 3.1, are designed for achieving higher compression rate and in many cases those enhancements result in noisier motion field. In compression tasks, commonly, this

motion estimation noise is dealt with in subsequent processing stages. In applications where motion estimation is the goal, additional processing or adjustments are required. Figure 3-2 depicts MPEG-4 motion field computed for a real-life video stream. Figures (b) is the motion field of the frame, presented in figure (a) and its subsequent one, computed using diamond search block matching algorithm [80] which commonly used in MPEG-4 codecs. The marking on figure (b), were added manually and represents the car's position. As one can see, while the car's motion is noticeable, there is a significant amount of additional false detection in the scene.

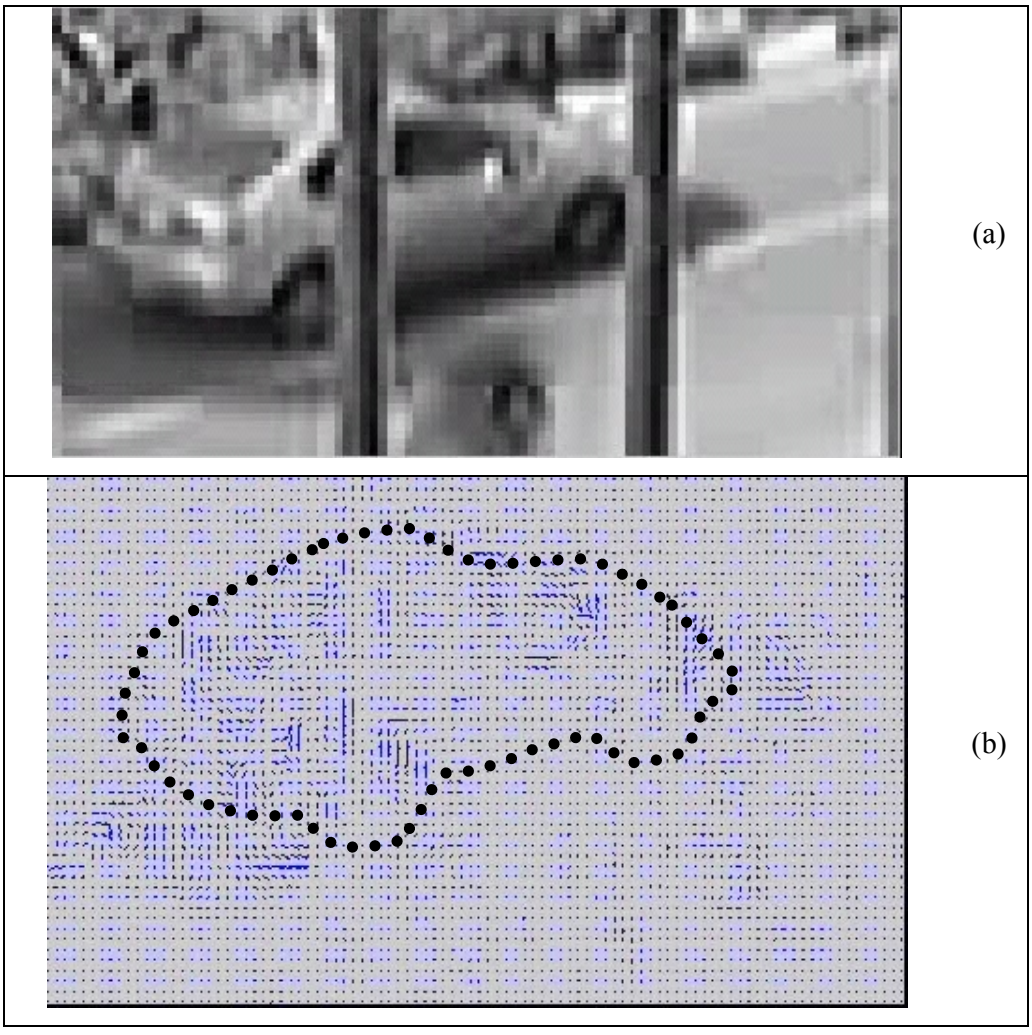


Figure 3-2 – MPEG-4 block matching based motion extraction. Figure (b) represent the motion field computed for the frame presented in figure (a) and its following one. The motion field was computed by the diamond search block matching algorithm, which is commonly used in MPEG-4 codecs. The dotted line was artificially placed on figure (b) to show the car's position on the motion field representation.

3.3. Optical Flow Methods

Estimation of local motion in a set of images, which is based upon local derivatives, is commonly referred to as optical flow [84-89]. The basic principle of optical flow computation can be described as follows. During time interval $\Delta\tau$ the intensity level at coordinates (ξ,υ) moves to $(\xi+\Delta\xi,\upsilon+\Delta\upsilon)$. It is assumed that the motion causes no changes in the intensity, and the changes may occur solely due to random factors such as additive signal independent white Gaussian noise that can be attributed to image sensor. Then, given image intensity measurements $I(\vec{\phi})$ and $I(\vec{\phi}+\Delta\vec{\phi})$, where $\Delta\vec{\phi}=(\Delta\xi,\Delta\upsilon,\Delta\tau)$, the statistically optimal maximum likelihood estimation of motion vector $(\Delta\xi,\Delta\upsilon)$ is found as the solution of the equation:

$$(\Delta\xi,\Delta\upsilon)\big|_{\tau=\tau_0}=\arg\min_{(\Delta\xi,\Delta\upsilon)}\iint_{(\xi,\upsilon)\in ARM}\left[I(\vec{\phi})-I(\vec{\phi}+\Delta\vec{\phi})\right]^2 d\xi d\upsilon \tag{3-3}$$

where ARM is the object Area of Rigid Motion centered at the point (ξ,υ). Within the accuracy of Taylor expansion of the image intensity function $I(\vec{\phi})$, equations (3-3) can be approximated by equation (3-4):

$$\begin{aligned}(\Delta\xi,\Delta\upsilon)\big|_{\tau=\tau_0}&=\arg\min_{(\Delta\xi,\Delta\upsilon)}\iint_{(\xi,\upsilon)\in ARM}\left[\frac{\partial I(\vec{\phi})}{\partial\xi}\Delta\xi+\frac{\partial I(\vec{\phi})}{\partial\upsilon}\Delta\upsilon+\frac{\partial I(\vec{\phi})}{\partial\tau}\Delta\tau\right]^2 d\xi d\upsilon=\\&=\arg\min_{(\Delta\xi,\Delta\upsilon)}\iint_{(\xi,\upsilon)\in ARM}\left(\dot{I}(\vec{\phi})\bullet\Delta\vec{\phi}\right)^2 d\xi d\upsilon\end{aligned} \tag{3-4}$$

where $\Delta\vec{\phi}$ is a vector of space-time shifts $(\Delta\xi;\Delta\upsilon;\Delta\tau)$ and $(.\bullet.)$ is a scalar (inner) vector product and $\dot{I}(\vec{\phi})$ is a vector of image intensity space-time derivatives:

$$\dot{I}(\vec{\phi})=\left(\frac{\partial}{\partial\xi}I(\vec{\phi}),\frac{\partial}{\partial\upsilon}I(\vec{\phi}),\frac{\partial}{\partial\tau}I(\vec{\phi})\right) \tag{3-5}$$

Two original algorithms of optical flow computation, Lucas-Kanade [84] and Horn-Schunck [85] implement modifications of equation (3-5). In the Lucas-Kanade algorithm, a weighting window function $W(\xi,\eta)$ is introduced:

$$(\Delta\xi,\Delta\upsilon)\big|_{\tau=\tau_0}=\arg\min_{(\Delta\xi,\Delta\upsilon)}\iint_{(\xi,\upsilon)\in ARM}W(\xi,\upsilon)\left(\dot{I}(\vec{\phi})\bullet\Delta\vec{\phi}\right)^2 d\xi d\upsilon \tag{3-6}$$

while in the Horn-Schunck algorithm an additional constrain on smoothness of the space shift vector is introduced and integration is extended to the entire image frame ($ImgFr$):

$$\left(\Delta\xi,\Delta\upsilon\right)\Big|_{\tau=\tau_0} = \underset{(\Delta\xi,\Delta\upsilon)}{\arg\min} \iint\limits_{(\xi,\upsilon)\in ImgFr} \left\{\left(\dot{I}\left(\vec{\phi}\right)\bullet\Delta\vec{\phi}\right)^2 + \lambda^2\nabla^2\right\} d\xi d\upsilon \tag{3-7}$$

where

$$\nabla^2 = \left(\frac{\partial\boldsymbol{\Delta}_\xi}{\partial\xi}\right)^2 + \left(\frac{\partial\boldsymbol{\Delta}_\upsilon}{\partial\upsilon}\right)^2 + \left(\frac{\partial\boldsymbol{\Delta}_\upsilon}{\partial\xi}\right)^2 + \left(\frac{\partial\boldsymbol{\Delta}_\upsilon}{\partial\eta}\right)^2 \tag{3-8}$$

and λ^2 is a weight parameter of the smoothness constraint ∇^2.

Since Horn-Schunck's and Lucas-Kanade's works, a number of modifications of these basic optical flow algorithms have been proposed aimed at the improvement of the accuracy and reliability of optical flow computations. Most of these efforts deal with the problem of local minima in the optimization process, possible intensity variations that cannot be described by the additive normal noise model, inaccuracies due to image sampling and improvement of numerical optimization schemes. The more recent method combining together some principles from the latest research of the optical flow algorithms [97,98,99,100,101,102] and reporting good results belongs to Brox et al [87]. This method exploits the intensity constancy and motion field smoothness assumption as in the Horn-Schunck method, and, in addition, introduces an assumption of constancy of image intensity gradient. Multi-scale minimization strategy is used as part of the entire iterative procedure.

The optimization functional used is:

$$\begin{aligned}\left(\Delta x,\Delta y,t\right) = \underset{(\Delta x,\Delta y)}{\arg\min} \iint\limits_{(\xi,\eta)\in ImgFr} \Big\{\Big(\Big|I\left(\xi+\Delta x,\eta+\Delta y,t+\Delta t\right)-I\left(\xi,\eta,t\right)\Big|^2 + \\ \gamma\Big|\nabla I\left(\xi+\Delta x,\eta+\Delta y,t+\Delta t\right)-\nabla I\left(\xi,\eta,t\right)\Big|^2\Big)+\varepsilon^2\Big]^{1/2} + \alpha\left(\nabla^2+\varepsilon^2\right)^{1/2}\Big\} d\xi d\eta\end{aligned} \tag{3-9}$$

where $\boldsymbol{\nabla^2}$ as in (5), $\boldsymbol{\varepsilon^2}$ is a user defined regularization constant and $\boldsymbol{\alpha,\gamma}$ - are weighting parameters. In addition, a multi-scale approach is applied with inter-scale updating as $\boldsymbol{\Delta^k = \Delta^{k-1} + d\Delta^k}$, where k is a scale index, $\boldsymbol{\Delta^{k-1} = (\Delta x,\Delta y)^{k-1}}$ is the displacement vector found in the previous (coarsest) scales and $\boldsymbol{d\Delta^k}$ is the displacement increment found in the current scale.

Some modifications of this algorithm dealing with motion segmentation were suggested in [86, 103]. An improvement of the numerical schemes and real time implementations is suggested in [104].

Figure 3-3 illustrates the optical flow method in real-life scenarios. Figure 3-3(b) depicts the motion field between the frame illustrated in Figure 3-3(a), and its subsequent frame, computed by the Horm and Shunck's optical flow method [85]. The motion field's x and y components are represented by the G and B components, respectively, of the RGB image of figure (b), the R component is zeroed. It is clearly seen that while the moving cars' motion is estimated, no noise motion is present in the scene.

3.4. Comparison Framework

Since introducing differential optical flow methods, a number of efforts to evaluate their performance in quantitative terms have been reported. For the evaluation, synthetic as well as real-life test images sequences with known optical flow "ground truth" were used.

In 1994, Barron et al. [88] presented the comparison overview of the optical flow methods described in [84,85,105,106,107,108,109,110,111]. The methods were compared using computer generated image "Translating Square" and real life test image sequences "Translating/Diverging Tree" with computer generated motion fields (methods of introducing artificial shifts were not reported in the paper). In addition, a "Yosemite" sequence suggested in [106], for which "ground truth" motion field was directly measured using digital terrain map of the "Yosemite" valley.

In 1995, M. Otte and H. Nagel suggested another test real life image sequence for method benchmarking [102]. The sequence was recorded with a calibrated camera fixed on the arm of a robot, which moves along a precisely defined 3D-trajectory. The original "ground truth" motion field was obtained from the camera calibration data.

In 1998, Liu et al. examined the accuracy/efficiency tradeoffs for algorithms described in [84,85,105,108,109,112,113,114] using the mentioned above "Translating/Diverging Tree" sequences for comparison [115].

In 1999, B. McCane et al. proposed, for generating test image sequences, using computer graphics methods of rendering of consecutive frame and calculation of the motion vectors using ray-tracing techniques to synthetic test objects [91]. The authors also consider an option of semi-manual measuring the "ground-truth" motion on certain real-life cartoon videos.

Most of recent publications on optical flow algorithms evaluate their performance using the aforementioned "Yosemite" sequence (see Figure 3-4), that is now provided

for common use (for instance, on *ftp://ftp.csd.uwo.ca/pub/vision*), though no data as to the accuracy of the "ground truth" motion field is available. Moreover, it was found recently, that the "ground truth" data provided for the "Yosemite" image sequence are, in fact, not always perfectly correct [103]. A common drawback of using this and, in fact, any other particular image sequence for comparison the optical flow methods is that it is not clear at all to what degree a particular image sequence represent all image sequences to which the methods are intended. Although researchers do appreciate this problem, (see, for instance [116]), no suggestion to solving this problem has been proposed.

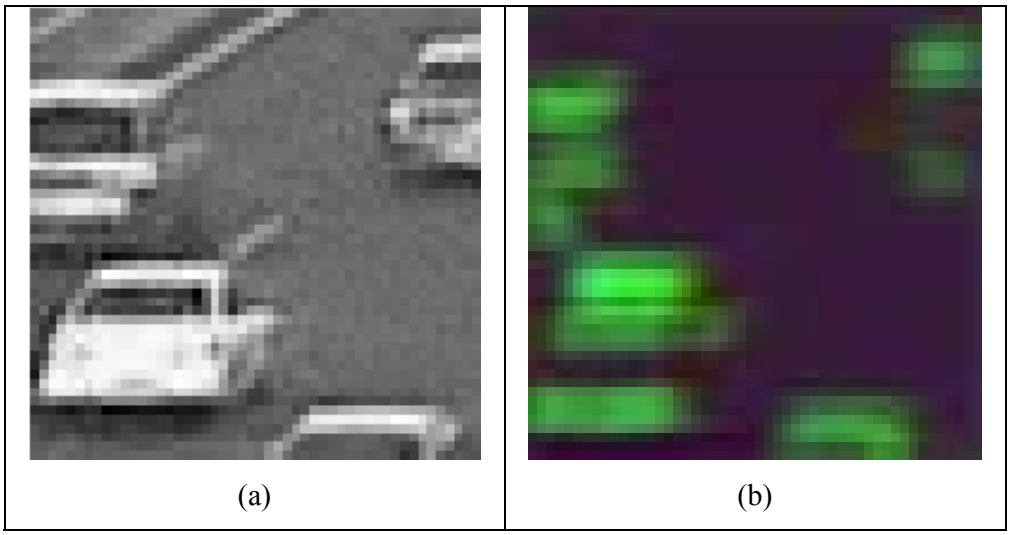


Figure 3-3 – Optical Flow motion estimation. Figure (b) is the motion field extracted from real-life traffic video using Horn and Shunck's optical flow method. Figure (a) illustrates the corresponding frame from the traffic video sequence. The motion field's x and y components are represented by the G and B components, respectively, of the RGB image of figure (b), the R component is zeroed.

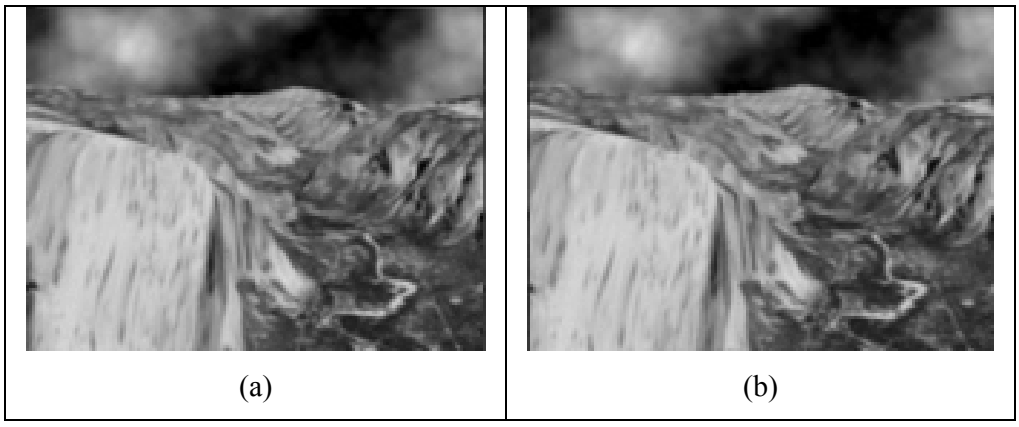


Figure 3-4 – The Yosemite ad-hoc optical-flow test sequence

In order to avoid those limitations and provide a firm base for comparison and evaluation of the accuracy of optical flow methods the following comparison framework was developed [117].

As derived from the above discussion, all optical flow methods rely on computing spatial derivatives of the image intensity function. As will be shown in the following section, the accuracy of numerical differentiation depends of the signal bandwidth. Therefore, in addition to the "Yosemite" image, a set of computer-generated pseudo-random images of 256x256 pixels with uniform spectrum within a circular fraction of the base band defined by the image sampling rate is suggested.

The method of generation of the images set is explained in Figure 3-5. A set of low-pass DCT domain filter masks is first generated. Each mask is specified by a parameter, α, the percentage of lower frequencies coefficients which are not filtered out. For example the mask which doesn't filter any of the image frequencies corresponds to effective bandwidth of α=1. The masks which filter out half and quarter of the higher-frequency coefficients correspond to α=0.5 and α=0.75 effective bandwidths respectively. This set of filters is applied on a binary pseudo-random noise synthetic image, which takes values 1 and -1 with equal probabilities. This image is referred to as the set's *seed-image*. Then the product of the seed image and low-pass filter mask is subjected to DCT transform thus producing images with uniform (in DCT domain) spectrum within a certain bandwidth defined by the filter mask. This is depicted in Eq. (3-10) and Figure 3-5. Figure 3-6 illustrates examples of such test images.

$$I_{test}(\alpha) = DCT\{I_{seed} \cdot LPF\ Mask(\alpha)\} \tag{3-10}$$

In order to generate sequences of shifted frames with known shifts, artificial shifts are introduced in the pseudo-random test images as well as in the "Yosemite" test image. First of all, image shifts by integer number of inter-pixel distance were used, which guarantied the absence of interpolation errors in the shifted test images. We also generated images with global sub inter-pixel distance shifts, using for this purpose discrete sinc-interpolation, which has been proved to be the perfect interpolation method for sampled data with finite number of samples [118]. Displacing source test images in this manner we obtain, for each specific test image, the consecutive frame with exactly known displacement vector for each pixel in the frame.

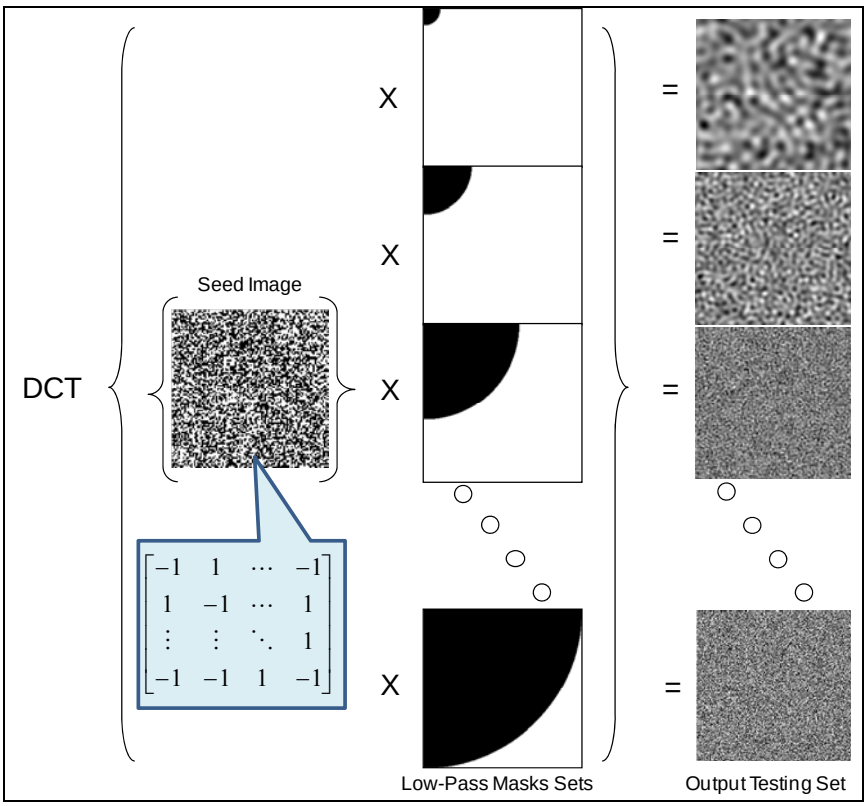


Figure 3-5 - The generation of the test images with known bandwidths.

3.5. Improved Optical Flow Accuracy through Precise Differentiation

An important issue in implementing and using differential optical flow methods is the accuracy of numerical estimation of the image derivatives. Although some authors did appreciate the importance of accurate computation of derivatives [116,119,120,121], no thorough quantitative analysis and comparison has been done.

J.K. Kearney et al. [121] presented analysis of error sources of the gradient based optical flow methods with local optimization such as, for instance, for instance, Lucas and Kanade method [84]). In particular, this work analyses the contribution of the "forward difference" differentiator, which evaluates the derivative as an inverse difference between the current signal sample and the next one, to the overall error of this family of methods. Simoncelli [116] and Elad et al. [119] attempted to find the best matching of the differentiation kernel to the low-pass pre-filter that is commonly used to preprocess images in differential optical flow. Brandt [120] investigates influence of various aspects of the method with local optimization to its performance: size and weight function of the optimization window, low pass pre-filter,

differentiators, etc. and, in particular, discusses the impact of three commonly used numerical differentiation techniques on the performance of this method.

In this section the use of DFT/DCT differentiation techniques for improving the accuracy of differential optical flow algorithms is suggested. This is justified by analytical and experimental results that demonstrate, in quantitative terms, the influence of the accuracy of numerical differentiation to the optical flow computation [117].

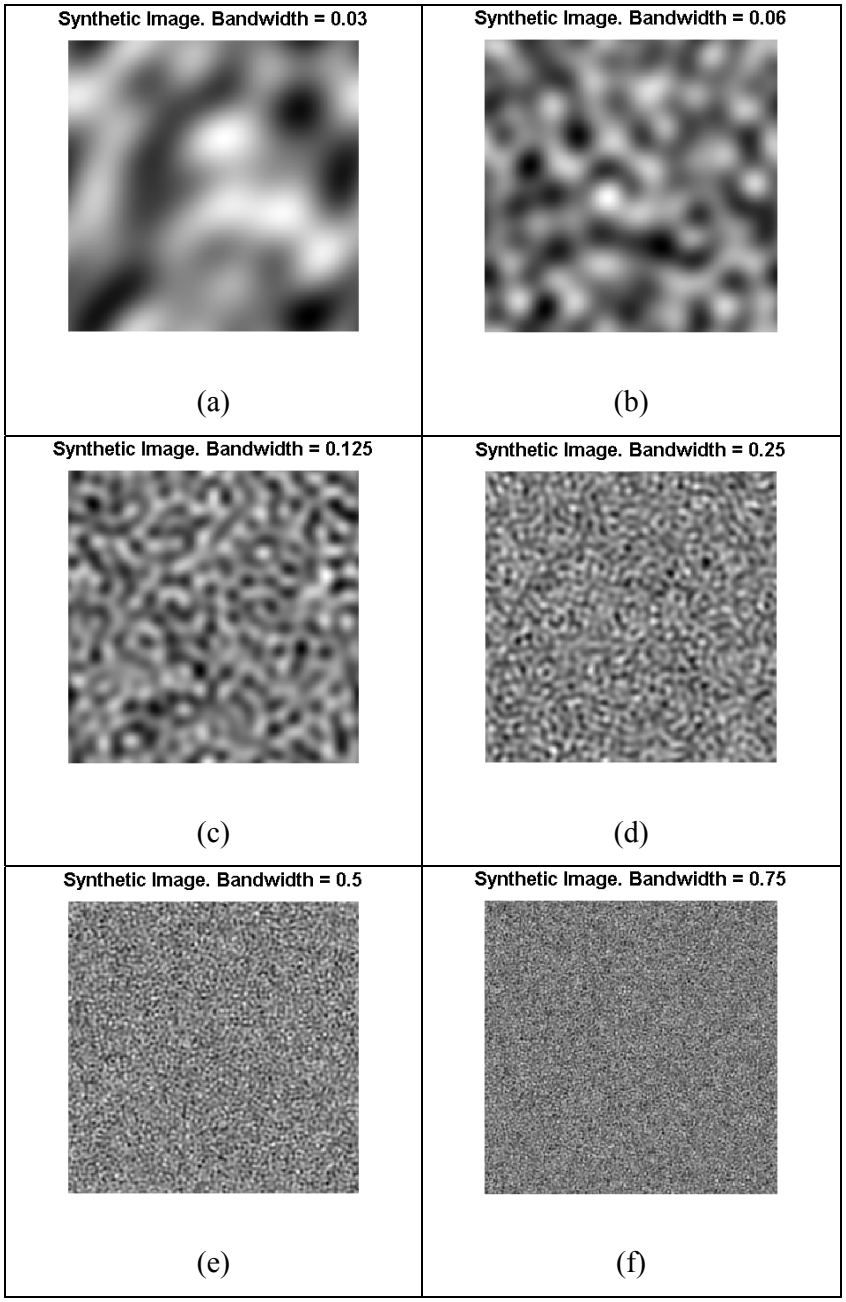


Figure 3-6 - Examples of test pseudo-random images with different bandwidth.

3.5.1. Differentiation Methods

For the evaluation of the differentiation methods impact on optical flow accuracy, 5 standard, widely used, differentiation methods are used: D1, D2, D4 [89, 122], the Simoncelli Kernel [116] as well as DFT/DCT based methods (DCT and DFT) are compared. The first 4 methods are implemented as digital convolutions with the following convolution kernels:

$$h^{D1} = [-1, 1] \tag{3-11}$$

$$h^{D2} = \left[-\frac{1}{2}, 0, \frac{1}{2}\right] \tag{3-12}$$

$$h^{D4} = \left[-\frac{1}{12}, \frac{2}{3}, 0, -\frac{2}{3}, \frac{1}{12}\right] \tag{3-13}$$

$$h^{Simoncelli} = [-0.108, -0.283, 0, 0.283, 0.108] \tag{3-14}$$

The PSF of the above numerical differentiation methods, D1, D2, D4 and Simoncelli are correspondingly given by:

$$\begin{aligned} \eta^{D1}(r) &= \exp\left(-i2\pi\frac{r}{N}\right) - 1 = \\ &= -2i\sin\left(\pi\frac{r}{N}\right)\exp\left(-i\pi\frac{r}{N}\right) \end{aligned} \tag{3-15}$$

$$\eta^{D2}(r) = -i\sin\left(2\pi\frac{r}{N}\right) \tag{3-16}$$

$$\eta^{D4}(r) = \frac{1}{6}i\left[\sin\left(4\pi\frac{r}{N}\right) - 8\sin\left(2\pi\frac{r}{N}\right)\right] \tag{3-17}$$

$$\eta^{Simoncelli}(r) = i\left[0.217\sin\left(4\pi\frac{r}{N}\right) + 0.566\sin\left(2\pi\frac{r}{N}\right)\right] \tag{3-18}$$

DFT/DCT based differentiation algorithms were implemented as filtering in DFT and in DCT domains, correspondingly [118]. DFT domain numerical differentiation algorithm is described by:

$$\left\{\dot{a}^{DFT}(x)\right\} = \mathrm{IFFT}_N\left(\left\{\eta^{diff}(r)\right\} \bullet \mathrm{FFT}\left(\left\{a(x)\right\}\right)\right) \tag{3-19}$$

where $\{a(x)\}$ and $\dot{a}^{DFT}(x)$ are N samples of a signal and its derivative, respectively, $(x = \{1,2,3,..., N\})$ and

$$\eta^{diff}(r)=\begin{cases} -\dfrac{i2\pi r}{N} & ,r=0,1,\ldots,\dfrac{N}{2}-1 \\ -\dfrac{\pi}{2} & ,r=N/2 \\ \dfrac{i2\pi(N-r)}{N} & ,r=\dfrac{N}{2}+1,\ldots,N-1 \end{cases} \tag{3-20}$$

$$\eta^{diff}(r)=\begin{cases} -\dfrac{i2\pi r}{N} & ,r=0,1,\ldots,\dfrac{(N-1)}{2}-1 \\ \dfrac{i2\pi(N-r)}{N} & ,r=\dfrac{(N+1)}{2},\ldots,N-1 \end{cases} \tag{3-21}$$

for even and odd values of N respectively.

DCT domain numerical differentiation algorithm is described by the equation:

$$\begin{aligned} \left\{\dot{a}^{DCT}(x)\right\} &= \mathrm{IDcST}\left(\left\{\eta^{diff}(r)\right\}\bullet \mathrm{DCT}\left(\{a(x)\}\right)\right)= \\ &= \frac{2\pi}{N\sqrt{2N}}\sum_{r=1}^{N-1} r\alpha^{DCT}(r)\cdot\sin\left(\pi\frac{k+1/2}{N}r\right) \end{aligned} \tag{3-22}$$

The absolute values for the frequency responses for the filters described by Eqs. (3-15)-(3-18) and Eqs. (3-20)-(3-22) are plotted for comparison in Figure 3-7. It is evident that the standard numerical differentiation methods (D1, D2, D4 and Simoncelli's) tend to produce significant errors for signal with bandwidths higher than 0.5 of the base-band defined by the signal sampling rate, while the DFT and DCT domains numerical differentiation methods implement exact differentiation of sampled signals. The advantage of the DCT–based algorithm with respect to DFT is that it is substantially less vulnerable to boundary effects, which are unavoidable in digital filtering because of the finite number of signal samples [118].

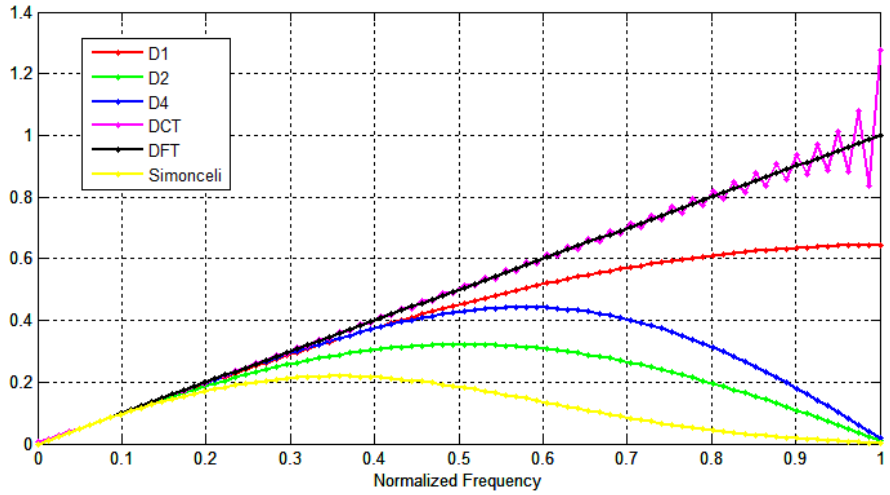


Figure 3-7 - Frequency responses of five numerical differentiation methods

3.5.2. Evaluation of the Error Associated with the Taylor Expansion

The lower bound of error in estimating optical flow does not depend on the methods of solving Eq. (3-4) and is determined by two factors: errors in evaluation of spatial derivatives and errors due to truncation of Taylor expansion. For the sake of simplicity, these errors are analyzed in 1-D using Discrete Fourier Transform signal representation. Let $\{a(x)\}$ and $\{\alpha(r)\}$ be the discrete signal of N samples and its discrete Fourier transform, correspondingly:

$$\alpha(r) = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} a(x) \exp\left(i2\pi \frac{x \cdot r}{N} \right) \tag{3-23}$$

$$a(x) = \frac{1}{\sqrt{N}} \sum_{r=0}^{N-1} \alpha(r) \exp\left(-i2\pi \frac{x \cdot r}{N} \right) \tag{3-24}$$

The finite, p-difference, $a(x+p) - a(x)$, then, can be computed via its DFT spectrum as

$$a(x+p) - a(x) = \frac{1}{\sqrt{N}} \sum_{r=0}^{N-1} \alpha(r) \left[\exp\left(-i2\pi \frac{p \cdot r}{N} \right) - 1 \right] \exp\left(-i2\pi \frac{x \cdot r}{N} \right) \tag{3-25}$$

Therefore frequency response of the finite difference operator is given by:

$$\eta^{fd}(r,p) = \exp\left(-i2\pi \frac{p \cdot r}{N} \right) - 1 = -2i \sin\left(\pi \frac{p \cdot r}{N} \right) \exp\left(-i\pi \frac{p \cdot r}{N} \right) \tag{3-26}$$

Without loss of generality let us assume that a signal sampled at time $\boldsymbol{t + \Delta t}$ is a p-shifted version of the signal sampled at time $\boldsymbol{t}$:

$$a_x^{t+\Delta t} = a_{x+p}^t \tag{3-27}$$

In optical flow algorithms, the finite difference is computed via its approximation through signal Taylor expansion, which suggests,

$$a_{x+p}^t - a_x^t = \dot{a}_x^t p \tag{3-28}$$

Representing Eq. (3-28) in the frequency domain results in:

$$DFT\left\{ a_{k+p}^t - a_k^t \right\} = DFT\left\{ \dot{a}_k^t p \right\} \tag{3-29}$$

The discrete Fourier transform of the difference $\left(a_{k+p}^t - a_k^t \right)$ is given by Eq. (3-25), $\dot{a}_x^t$ is given by Eq. (3-19). Now one can find energy spectrum of the Taylor expansion approximation error:

$$\sigma^2(\Delta) = \left|\eta^{fd}(r,\Delta) - \eta^{Taylor}(r,\Delta)\right|^2 = \\ = \left|2i\sin\left(\pi\frac{\Delta\cdot r}{N}\right)\exp\left(-i\pi\frac{\Delta\cdot r}{N}\right) + \Delta\cdot\eta^{diff}(r)\right|^2 \quad (3\text{-}30)$$

By virtue of Parceval's theorem for DFT, cumulative summation of this spectrum will give, for a given *p*-shift, the standard deviation of the Taylor approximation error as a function of the signal bandwidth. This will allow comparing, for different signal shifts, different differentiation methods in terms of the lower bound of error of optical flow computation.

Figure 3-8 shows the estimation error for different shifts as a function of the signal's bandwidths. As can be seen the optical flow computation using DFT/DCT differentiation outperforms all lower order numerical differentiators, especially for signals with broad bandwidth. For shifts larger than 0.4 signal inter-pixel distance, error's nominal magnitudes are substantially larger. For larger shifts, lower order differentiators have lower Taylor approximation error, although, by its nominal magnitude, the errors increase substantially. Not surprisingly, for one-pixel shift approximation error for zero-order D1 differentiator is zero, as can be seen in Figure 3-8(f). The conclusion from this comparison is that the differentiation-based optical flow computation methods can work with reasonable accuracy only in the range of small displacements, which firmly motivates using their multi-scale versions.

3.5.3. Comparison Results

For the comparison purpose, multi-scale versions of two basic optical flow algorithms have been selected: (i) Lucas-Kanade (L&K) [84] and (ii) Horn-Schunck (H&S) [85] in which a multi-scale (multi-grid) framework was imbedded in order to enable measuring large scale shifts (up to 1-3 pixels). Additionally, the comparison was held with respect to one of the more recent methods presented by Brox et al (Brox) [87]. To this end, in-house implementation was used as the original implementation of the algorithm is not available.

The following numerical parameters were opted for the best algorithm performance over the above-mentioned set of test images. For all algorithms there are 10 scale levels with reduction factor of 0.8. The reduction was implemented using bi-cubical interpolation. The optimization window size for the L&K method was selected to be (5x5) pixels. The smoothness parameter λ, for the H&S method, was

set to 10^3 and the number of Gauss-Seidel iterations to 50. 10 inner fixed point iterations and 20 successive over-relaxation-method iterations were used within the Brox method. The smoothness parameter, α, and the gradient constancy weight, γ, in Brox's notation, were set to 80 and 100 respectively [87]. Image frame pre-filtering used in all methods was implemented with commonly used setting of the sigma-parameter set to 1 ($\sigma = 1$) [87-89].

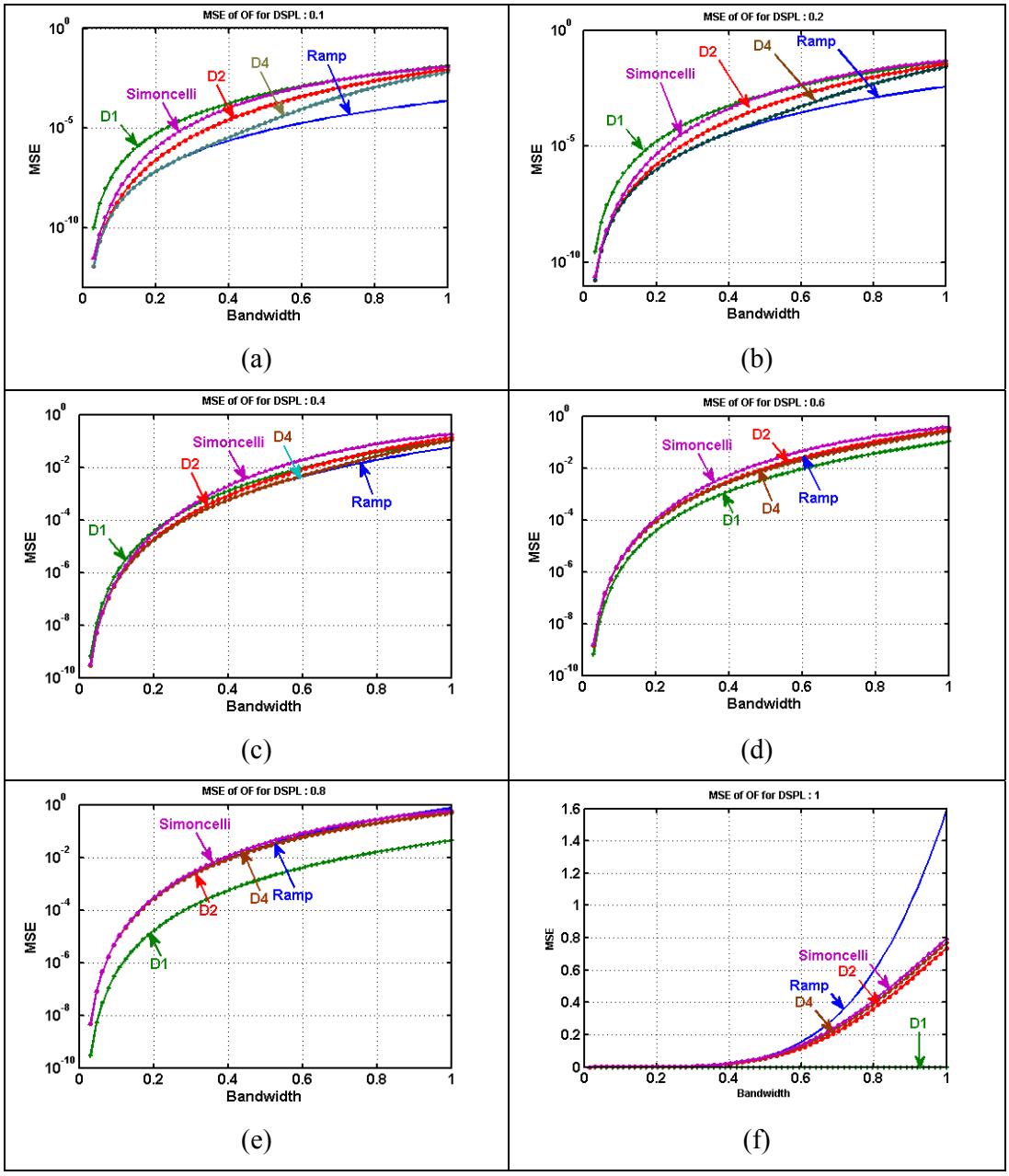


Figure 3-8 – Lower bounds of shift estimation as a function of signal bandwidth for 1D optical flow computed using different differentiation methods for different shifts.

In all set of experiments, global as well as local shifts were used and statistics of optical flow coordinate shift evaluation errors (distribution density histograms, standard deviation and mean values) were found over the set of all image's pixels. The errors were computed as the Euclidean difference between the pixels' computed shifts, $(\bar{x}, \bar{y})$, found by the various optical flow methods, and the known ones, (x, y).

$$e = \sqrt{(\bar{x} - x)^2 + (\bar{y} - y)^2} \tag{3-31}$$

In order to avoid boundary effects in differentiation by means of DFT based algorithm, the errors were analyzed within "safe" internal area of images separated from image borders by margins of 32 pixels.

For evaluation, the different differentiation methods were embedded into the optical flow schemes. As a test bed, pseudo-random images with effective bandwidth of 0.75 (as depicted in Figure 3-6(f)), shifted in 0.5, 1, 1.5 and 2 pixels were used. The results are summarized in Figure 3-9 for the L&K (figure a) and the H&S (figure b) methods. As one can see, for both optical flow techniques the DFT and DCT differentiation approaches outperform the others.

The accuracy gain factor, *G,* is defined as the ratio of the errors' standard deviations, for a given optical flow technique, when the derivatives are computed with different derivations methods:

$$Gain\,Factor\big|_{OpticalFlow} = \frac{\underset{x,y\in\Omega}{std}\left(e(x,y)_{Diff\,(1)}\right)}{\underset{x,y\in\Omega}{std}\left(e(x,y)_{Diff\,(2)}\right)} \tag{3-32}$$

where $\underset{x,y\in\Omega}{std}\left(e(x,y)_{Diff\,(1)}\right)$ and $\underset{x,y\in\Omega}{std}\left(e(x,y)_{Diff\,(2)}\right)$ are the errors' standard deviations , in neighborhood Ω, for a given *OpticalFlow* method embedding the *Diff(1)* and *Diff(2)* differentiation methods. Table 3-1 presents the accuracy gain factor for D2/DCT and D4/DCT embedded in the L&K and H&S optical flow techniques. The table suggests that a significant improvement is achieved when embedding the DCT exact differentiation method.

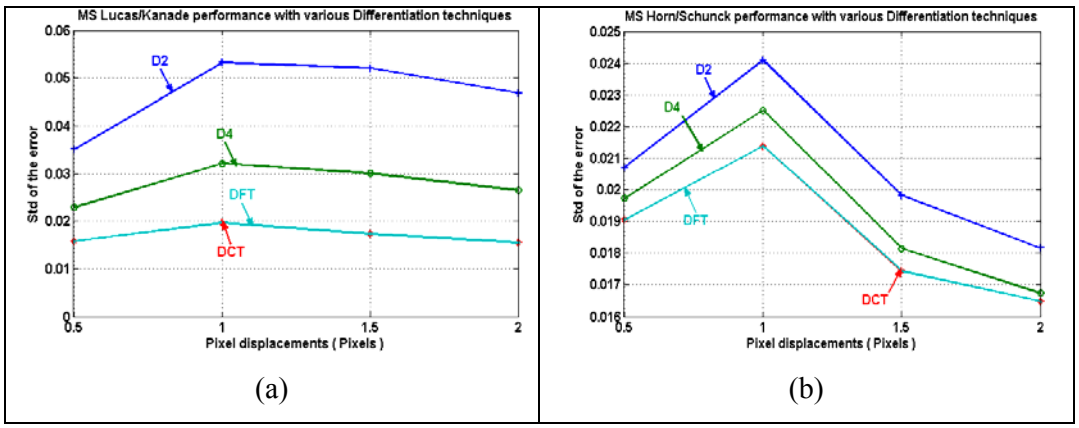


Figure 3-9 - Standard deviations of coordinate shift error, for different displacements (in the units of inter-pixel distance), of the L&K (figure a) and H&S (figure b) optical flow methods embedding the different differentiation methods.

Table 3-1 – Accuracy Gain Factor for D2/DCT and D4/DCT embedded into the L&K and H&S optical flow methods for different displacements.

Optical Flow Method	Differentiation Method	Displacement (pixels)			
		0.5	1	1.5	2
L&K	D2	2.22	2.71	3.01	3.02
	D4	1.45	1.64	1.74	1.71
H&S	D2	1.09	1.13	1.14	1.1
	D4	1.04	1.05	1.04	1.02

The effect of the different derivation methods on optical-flow techniques, with respect to the test-images' bandwidths is depicted in Figure 3-10, where the error's standard deviations are presented for test images with bandwidths of 0.03, 0.06, 0.125, 0.25, 0.5 and 0.75, shifted in one pixel, for different differentiation methods and the two selected optical flow algorithms, L&K (a) and H&S (b). The gain factors, D2/DCT and D4/DCT, embedded into the L&K and H&S schemes, for the various bandwidths test images are given in Table 3-2. The plots and the table suggest that a substantial gain in accuracy can be obtained, when using better differentiation techniques, especially when the processed images contain substantial high frequency content.

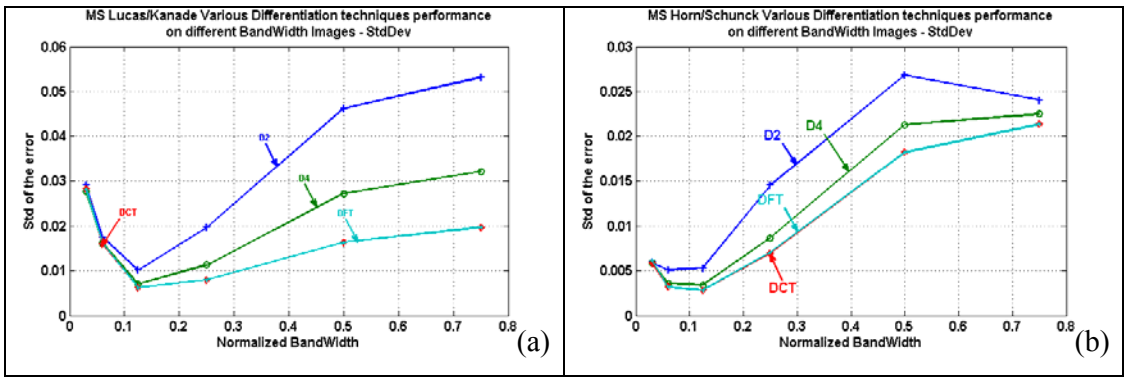


Figure 3-10 - Standard deviations of coordinate shift error of the L&K (figure a) and H&S (figure b) for pseudo-random test images, with ranging bandwidths (0.03, 0.06, 0.125, 0.5 and 0.75), applying shifts of one pixel.

Table 3-2 - Accuracy Gain Factor for D2/DCT and D4/DCT embedded into the L&K and H&S optical flow methods for different images bandwidths.

Optical Flow Method	Differentiation Method	Image Bandwidth					
		0.03	0.06	0.125	0.25	0.5	0.75
L&K	D2	1.03	1.1	1.62	2.47	2.83	2.71
	D4	1	1.02	1.13	1.42	1.66	1.64
H&S	D2	1.01	1.58	1.89	2.1	1.47	1.13
	D4	1	1.1	1.21	1.24	1.17	1.05

The quantitative data on the standard deviation of the optical flow estimation error obtained on the "Yosemite" test image with global shift by one inter-pixel distance in each axis are summarized, for L&K and H&S in Table 3. These data also demonstrate improvement in the accuracy of optical flow estimation achieved with more accurate differentiation techniques.

Table 3-3 – Error Standard deviation for the Yosemite test sequence for the L&K and H&S optical flow methods, embedding the D2, D4 and the DCT differentiation methods.

Optical Flow Method	Differentiation Method		
	D2	D4	DCT
L&K	0.126	0.084	0.061
H&S	0.057	0.046	0.040

Figure 3-11 illustrates the impact of the use of the DFT/DCT exact derivatives impact on the Brox optical flow method using the Yosemite test image (see Figure 3-4). Figures (a) and (b) illustrate the *pixel-wise* gain factor which is given by:

$$Gain\ Factor(x,y)\Big|_{OpticalFlow} = \frac{e(x,y)_{Diff(1)}}{e(x,y)_{Diff(2)}} \tag{3-33}$$

Figure 3-11(c) presents local image intensities standard deviation (over 7x7 window). Higher spatial local frequencies manifest themselves in higher standard deviation values, which correspond to brighter pixels. Evaluating figures (a) and (b) with respect to figurte (c), clearly shows that, generally, less accurate differentiation methods tend to produce higher optical flow estimation errors where image contain high frequencies. However, in some places, specifically, where the errors are small, this gain factor is less then 1, which can be attributed to the phenomenon, described in Sect. 3.5.2, that lower order derivation methods can give, for sufficiently large shifts, lower Taylor series expansion approximation error than higher order methods.

The experiments show that the Brox algorithm provides the best accuracy of optical flow computation. With that, its accuracy, paradoxically enough does not practically depend on the differentiation method used. This can be explained by the very high value of its smoothness-constrained weight compared to the gray-level-constancy constraint, in which the precise differentiation plays an important role. For greater gray-constancy weights, however, the algorithm doesn't always converge. This implies that the method is applicable for very smooth motion fields only [86].

The results demonstrate that the performance of L&K, H&S methods can be improved substantially provided accurate computation of spatial derivatives, especially for images characterized by high frequency content. Those conclusions, when applied on the L&K algorithm are significant. Since this algorithm is computationally light, easy to implement and more stable than methods implementing global iterative optimization procedures, like the H&S and the Brox methods, when implemented with the precise differentiation, it becomes an attractive option for optical flow computation.

3.6. Summary

Efficient real motion extraction is the first requirement in order to secure successful video restoration, resolution enhancement and scene reasoning. There are two classes of motion estimation techniques: block matching and optical flow methods. The first is based on dividing the current frame into a matrix of ‘macro blocks’ that are then compared with corresponding block and its adjacent neighbors in the previous frame. The latter is based upon local derivatives. In order to allow proper

evaluation of the various methods a comparison framework was suggested. The framework assumes using a set of test images with known bandwidths and mutual translations. Additionally, an improvement, through exact DFT/DCT-based derivatives, was suggested for optical flow methods. The gain in accuracy of the optical-flow methods was illustrated by embedding the exact derivatives into the L&K, H&S and Brox optical flow algorithms.

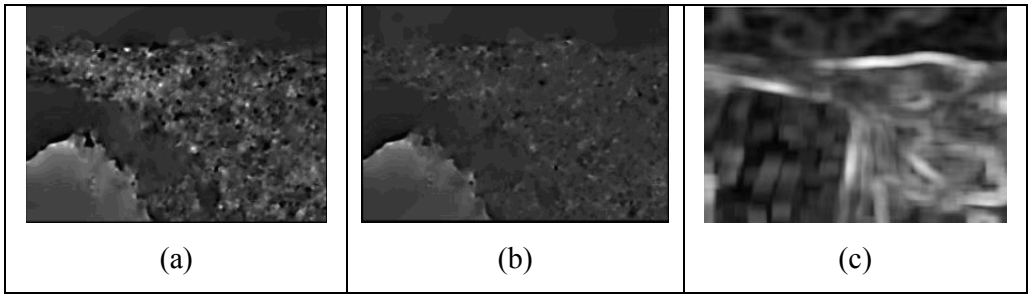


Figure 3-11 - Figures (a) and (b) illustrate the D2/DCT and the D4/DCT pixel-wise accuracy gain factors for the L&K and H&S methods respectively for global shift of one inter-pixel distance in each axis. Figure (c) depicts the 'Yosemite' test image's local intensity levels standard deviation computed over (7x7) window.

4. Video Stabilization

In numerous applications, one would like to locate and track real moving objects in the scene. However, this task becomes difficult, when the scene contains motion not only due to the real moving objects, but also due to camera or turbulent light propagation medium motion. This chapter describes a real-time image stabilization algorithm. To preserve real motion in the scene, the scene is segmented to moving objects and static areas and the compensation is applied only to the static areas of images. This is achieved through a statistical analysis of the motion vectors' magnitudes and angles, which results in formation of magnitude and angle motion segmentation driven masks. Performance evaluation of the algorithm is performed by simulation software which induces turbulence-like degradations on the input videos.

4.1. Channel Characterization and Processing Principles

As was described in Sect. 1.2, image processing based turbulence compensation algorithm that preserves real moving objects, without prior knowledge about the scene observed, were suggested in [21,22,23]. However, these methods do not comply with real-time constrains.

The real-time algorithm, suggested here, consists of three processing stages, illustrated in Figure 4-1:

(i) Estimation of the reference frames.

(ii) Determination of the motion vectors for all pixels in image frames and motion vector analysis for real motion extraction.

(iii) Generation of stabilized frames while preserving real motion in the scene.

The processing stages are thoroughly described in sections 4.2, 4.3 and 4.4, respectively.

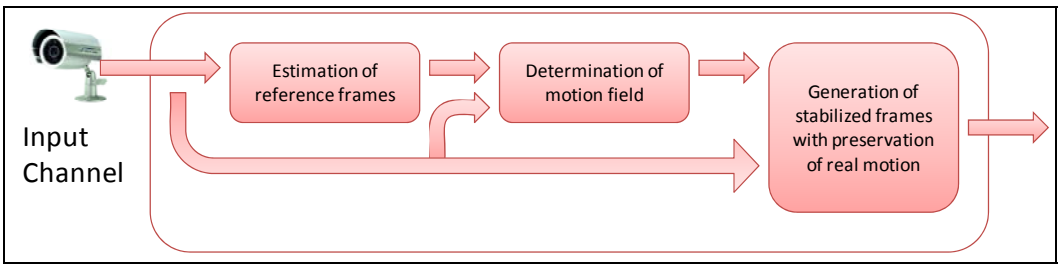


Figure 4-1 - Flow diagram of video processing in visual range channel

4.2. Estimation of the Reference Frames

In practice, in order to generate a stabilized sequence, one has to, for each time window, (i) take the current frame as the reference; (ii) compute motion vectors for each frame in the time window with respect to this reference frame; (iii) find pixel-wise means for motion vectors; (iv) resample pixels of each frame to the positions defined be the means. However this is applicable only in the absence of true moving objects that will create outliers in evaluation of motion.

The suggested solution generates the reference image using a pixel-wise rank filtering in a temporal sliding window. Specifically, temporal median filter is used for estimation of the stable scene [61], although other rank filters such as alpha-trimmed mean can also be considered. The size of the window, or the number of frames $\boldsymbol{N}$, over which the temporal rank filtering is carried out, is determined by the atmospheric turbulence correlation over the time domain, meaning that the longer the turbulence effect's correlation, the larger the temporal sliding window. In addition, it is important that the number of the images be sufficient enough to efficiently remove the moving objects from the stable scene estimation. Figure 4-2 illustrates such an estimation of a stable scene from a real life turbulent video. Figure 4-2(a) presents a single frame taken from a turbulent distorted sequence (the entire sequence can be downloaded from [123]). Figure (b) presents the estimation of the stable scene calculated by temporal median over 117 frames. One can clearly see that chaotic geometrical distortions in (a) are removed in the stable estimation (b).

4.3. Real Motion Extraction

In order to avoid, in course of the turbulence compensation process, compensation of real motion, pixels that represent real moving objects must be extracted from the observed frames. A following two-stage decision-making

algorithm for moving object extraction meets requirements of efficient real-time computation.

The first step is aimed at extracting areas, such as background, that are most certainly stationary and can be segmented in a very simple and fast way. In most cases, a great portion of the stable parts in the scene will be extracted at this stage. The rest of the pixels are dealt with at the second step.

The second step improves extraction accuracy at the expense of higher computational complexity, but it handles a substantially smaller portion of the pixels. This stage uses, for motion segmentation, computing and statistical analysis of optical flow [124,125,126,127].

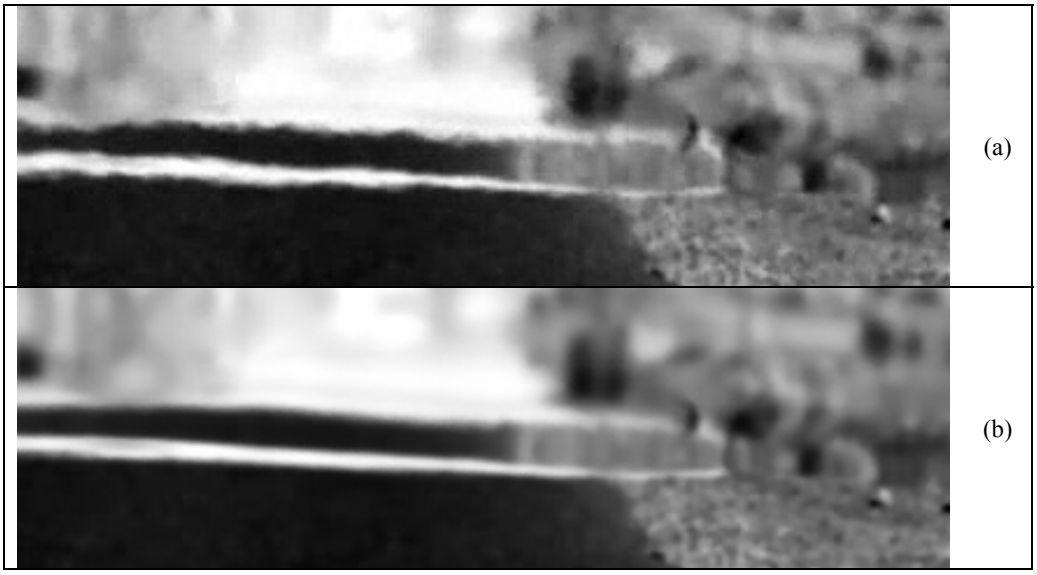


Figure 4-2 - Temporal Median Rank Filtering for estimation of the stable scene: (a) - a sample frame taken from a turbulent distorted video; (b) – the corresponding stable scene estimation.

4.3.1. Real Motion Extraction – stage I

At this phase, the gray-level difference between running value of each pixel of the incoming frame and its temporal median (Distance From Median, $DFM(\vec{p})$),is calculated as "real-motion measure" ((4-1)).

$$DFM(\vec{p}) = \left| I(\vec{p}) - \bar{I}(\vec{p}|_{t\in\Omega}) \right| \tag{4-1}$$

where $\bar{I}(\vec{p})$is the temporal median computed over a temporal window (Ω) centered at *t*, the current processed frame:

$$\bar{I}(x,y) = \underset{t\in\Omega}{MED}\{I(\vec{p})\} \tag{4-2}$$

If the distance $DFM(\vec{p})$ is below a given pre-defined threshold, the pixel is considered to be of a stationary object. All other pixels that are not resolved at this stage are processed at the next phase.

Figure 4-3 illustrates real-motion extraction using the above distance measure from the reference frame. Figure 4-3(a) is a single frame taken from a turbulent degraded sequence. The frame presented in Figure 4-3(b) is the stable scene median estimation. Applying a difference threshold of 10 gray-levels, on the pixel-wise absolute difference between figures (a) and (b), results in (c) (Higher difference values are printed in white), which, as depicted in Figure 4-3(c), is sufficient for filtering most stationary areas effectively. Evaluating figure (c), the car is detected as real moving object, and 75% of the pixels in the frame are tagged as stationary and will not, therefore, be further processed. However, some areas on the background that contain no motion are also tagged as real moving objects; this movement is due to turbulence. Those areas are dealt with at the following stage.

4.3.2. Real Motion Extraction – stage II

The second motion extraction stage uses more sophisticated optical flow analysis methods in order to achieve better discrimination accuracy. The mapping of one turbulent image to a stable one can be obtained by registering a spatial neighborhood surrounding each pixel in the image in the reference image. Such a registration can be implemented using different motion estimation methods (see Ch. 3.).

Let $I(\vec{p})$ be a turbulent source image frame and $\bar{I}(\vec{p})$ be the corresponding reference image, the vectorial difference between the pixel's location in the original image and its location in the reference image be the motion vector $\{\Delta x = x - \bar{x}; \Delta y = y - \bar{y}\}$. For the subsequent processing stages, the translation vector is presented in polar coordinates, hence magnitude and angle $\{M(\vec{p}), \theta(\vec{p})\}$ of the motion vector. Having the motion field, one can discriminate real motion from turbulent one through a statistical analysis of the Magnitude $\{M(\vec{p})\}$ and Angle $\{\theta(\vec{p})\}$ components of the motion field.

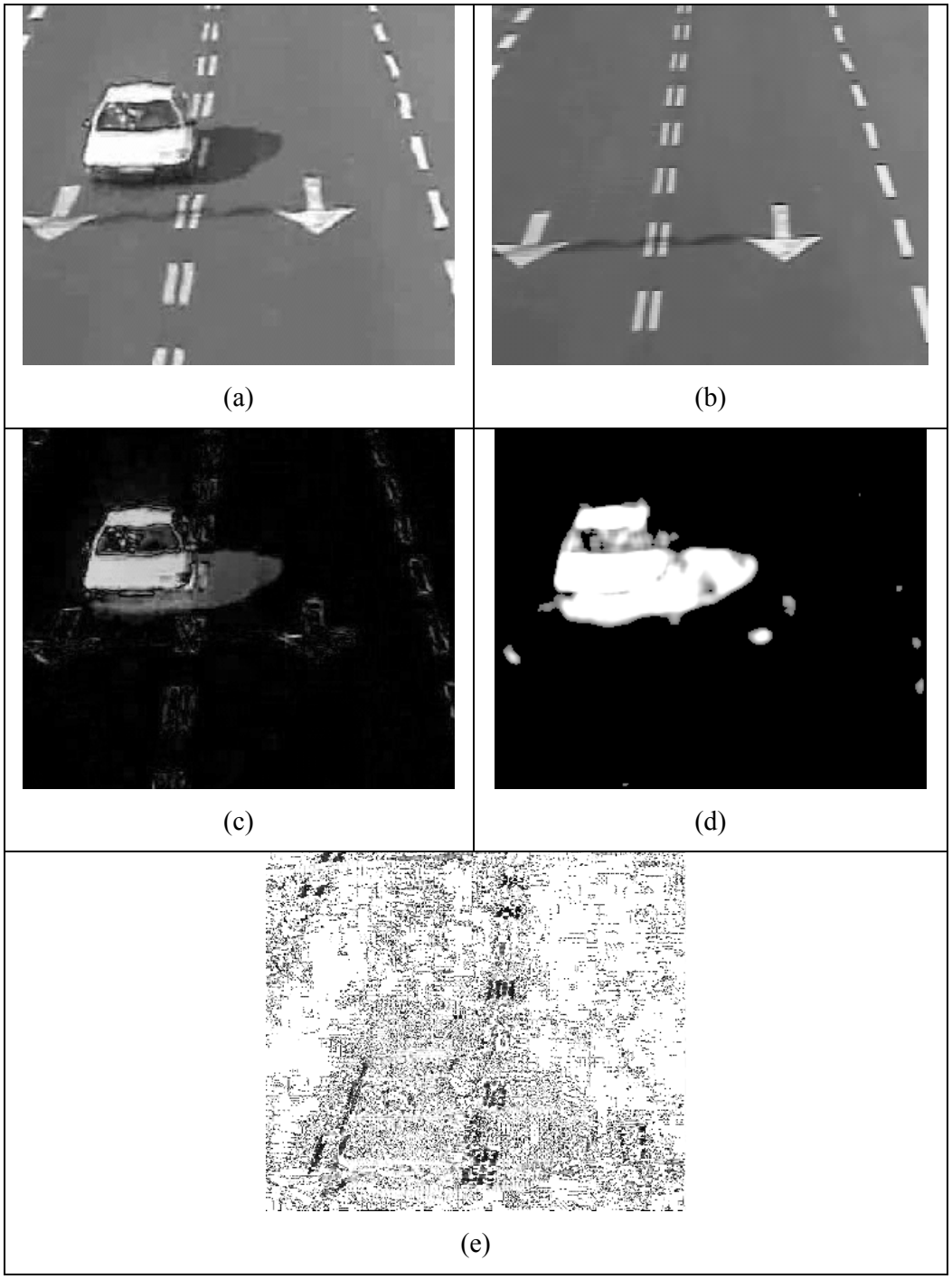


Figure 4-3 - Magnitude Driven Mask (MDM) and Angle Driven Mask (ADM). Figure (a) presents a single frame extracted from real-life turbulent degraded sequence. Figure (b) is the estimated stable scene. Figure (c), (d) and (e) present real-motion extraction by applying DFM, MDM and ADM respectively.

4.3.2.1. Real Motion vs. Turbulence Caused Motion Discrimination through Magnitude's Distribution Cluster Analysis

Cluster analysis of the Magnitude distribution function for all (*x, y*), in a particular frame *t*, allows separating two types of motion amplitudes: small and irregular and large and regular. The first is associated with small movements caused by turbulence. The latter corresponds to movements caused by real motion. At the result, each pixel in the frame is assigned with a certainty grade between 0 and 1. This "Magnitude Driven Mask" ($MDM(\vec{p})$) characterizes magnitude based likelihood that particular pixels belong to objects in a real motion. Figure 4-4 presents a graph of the certainty as a function of the Motion vectors' magnitudes. Small Motion Vectors' magnitudes correspond to turbulent motion, while large magnitudes correspond to real-motion. The intermediate levels comprise motion vectors' magnitude upon which concise decision can not be made. The magnitudes' thresholds presented as $\boldsymbol{T_L}$ and $\boldsymbol{T_H}$ are application dependent and can be set by the user. In some applications, where the conditions are maintained, the system can set those thresholds automatically. Figure 4-3(d) presents the MDM extracted from the frame presented in Figure 4-3(a). Pixels where real-motion was detected are marked in white. In this example $\boldsymbol{T_L}$ and $\boldsymbol{T_H}$ were set empirically and are 2 and 4 pixels respectively.

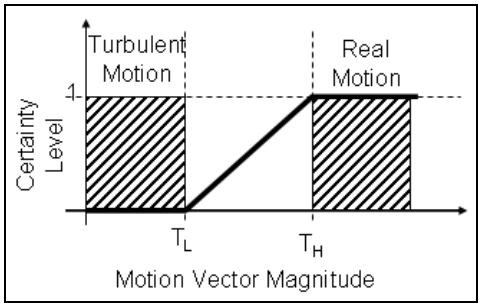


Figure 4-4 - Magnitude Driven Mask (MDM). MDM certainty level as a function of the motion vector's magnitude

4.3.2.2. Real Motion vs. Turbulence Caused Motion Discrimination through Motion Field's Angle Distribution

Pixel's motion discrimination through motion vector angle distribution is achieved by means of statistical filtering of the angle component motion field. For each pixel, its neighborhood's angle's standard deviation is computed. Turbulent motion has chaotic directions. Therefore, a motion field, in a small spatial

neighborhood, distorted by turbulence, has considerably large angular standard deviation. Real motion, on the other hand, has strong regularity in its direction and therefore its angles' standard deviation value over a local neighborhood will be relatively small. Homogeneous background areas contain no motion. Therefore the standard deviation of the zero motion vectors will be zero as well. The neighborhood size, in which the pixel's angular standard deviation is computed, should be large enough to make angle based discrimination of turbulent from real motion possible, and as small as possible to meet the terms of real-time computing. In experiments with real database it was found that neighborhood's size of 11x11-15x15 present a reasonable compromise.

As it is illustrated in the graph presented in Figure 4-5, each pixel is assigned with "Angle Driven Mask" ($ADM(\vec{p})$), which presents an angle distribution based likelihood that this pixel belongs to an object in a real motion. Both turbulent and background areas should be regarded as stable. This means that real moving objects have a bounded angular standard deviation. T_L and T_H are the decision boundaries. Pixels with angular standard deviation smaller than T_L or higher than T_H are regarded as stationary. Those values are set by the observer.

Figure 4-3(e) presents the ADM extracted from the frame shown in (a), with low standard deviation values displayed with brighter pixels. One can see that background areas are tagged in white, hence contain low standard deviations. The signs on the road, which suffer most from turbulent in the acquired scene, have darker values, hence, high angular standard deviations. $\boldsymbol{T_L}$ and $\boldsymbol{T_H}$ were empirically assigned with $\frac{\pi}{3}$ and $\frac{\pi}{6}$ respectively.

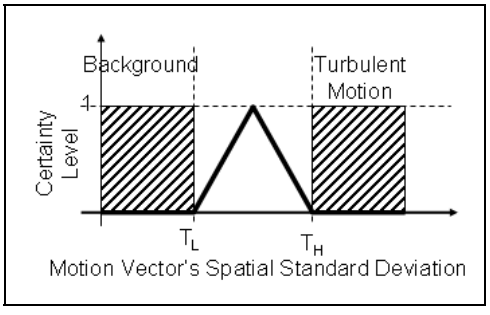


Figure 4-5 - Angle Driven Mask (ADM). ADM certainty level as a function of the motion vector's local spatial standard deviation

4.3.2.3. Real Motion Separation Mask

Having both the $MDM(\vec{p})$ and $ADM(\vec{p})$, a combined Real Motion Separation Mask ($\mathrm{RMSM}(\vec{p})$) is formed as following:

$$RMSM(\vec{p}) = \begin{cases} ADM(\vec{p}), & \left|ADM(\vec{p}) - \frac{1}{2}\right| > \left|MDM(\vec{p}) - \frac{1}{2}\right| \\ MDM(\vec{p}), & otherwise \end{cases} \qquad (4\text{-}3)$$

The MDM and ADM are certainty measures ranging from 0 (turbulent motion) to 1 (real motion). Equation (4-3) implies that the ADM measure is more concise than the MDM when the term $\left|ADM(\vec{p}) - \frac{1}{2}\right|$ has a higher value than $\left|MDM(\vec{p}) - \frac{1}{2}\right|$. In this case the ADM measure will be used; otherwise the MDM value will be applied.

4.4. Generation of stable frames

Based on the notations derived in the previous sections, the output frame $F_{(x,y,t)}$ is given by:

$$F(\vec{p}) = \bar{I}(\vec{p}) \bullet \{1 - [RMSM(\vec{p}) \bullet (1 - DFM(\vec{p}))]\} + I_{(x,y,t)} \bullet [RMSM(\vec{p}) \bullet (1 - DFM(\vec{p}))] \qquad (4\text{-}4)$$

where $F(\vec{p})$ is the output stabilized image, "$\bullet$" denotes element-wise matrix multiplication, ***DFM*** is the mask derived from the distance from the temporal median measure, as described in Sect. 4.3.1 and ***RMSM*** is the Real Motion Separation Mask detailed in Sect. 4.3.2.

4.5. Simulation and Results

For evaluating the suggested method, a testing visual database has been composed [123]. The database contains several typical thermal and visual real-life video sequences as well as synthetic ones. The synthetic sequences are generated from real-life sequences containing real-movement with no turbulent motion. Turbulent-like degradations are induced using computer software. The process of generating images with turbulence-like distortions is outlined in Sect. 4.5.1 and the image stabilization algorithm quantitative and qualitative results are respectively detailed in sections 4.5.2 and 4.5.3. Throughout this Sect., the dense-map was computed using the Horn and Schunck's optical flow method [85].

4.5.1. Turbulence Simulation Software

According the theory outlined in sections 1.2 and 2.2, the turbulence results in spatially and temporally correlated random geometrical distortions of the acquired image. In order to generate local correlated shifts for each pixel in the input image, a band limited correlated distortions field is generated. Figure 4-6 illustrates the process. For each direction, X and Y, a 3D random white noise 'cube' is first created. Then, a 3D Low-Pass filter is applied to the generated noise to produce spatial-temporal band-limited noise that defines a displacement field for each spatial coordinate. For shifting each pixel according to the displacement field, sliding window discrete sinc-interpolation algorithm [118] is applied.

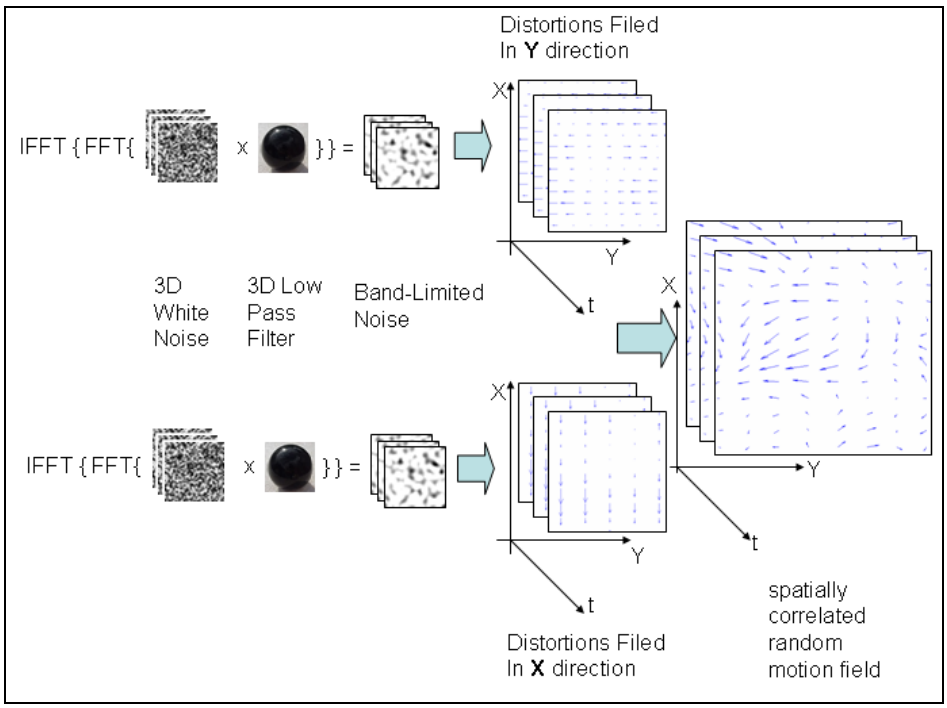


Figure 4-6 – Flow chart of the algorithm for generation of spatially and temporally correlated motion fields

4.5.2. Turbulence Compensation - Quantitative Evaluation

The synthesized test sequence is a turbulent free video stream with real motion in which a turbulent like degradation is induced. The compensation process aims at removing the turbulent motion while retaining real moving objects. Figure 4-7 illustrates the compensation results achieved on the test sequence. Figure 4-7(a) shows a frame of the initial non-distorted sequence, while figure (b) is the

corresponding frame from the sequence after applying turbulence-like degradations. Figure 4-7(c) shows the compensated corresponding frame of figure (b). Comparing Figures (a), (b) and (c), one can see that while the vehicle, as a real moving object, is retained, the turbulence distortions visible, for instance, on road markings in figure (b) of the background are compensated.

Quantitatively, the turbulence compensation quality can be evaluated using the peak signal to compensation error ratio (PSNR), as defined by (4-5):

$$PSNR = 10\log_{10}\left(\frac{(\max\{I\})^2}{MSE}\right) \tag{4-5}$$

where the MSE is the mean squared error computed over all image pixels and $\max\{I\}$ is the maximum value can be assigned to a pixel, hence 255.

In the given example, the *PSNR gives a high grade of 34.9 dB*. Typical values for the PSNR in images are between 30 and 40 dB, where higher values than 34 are consider as good quality [128].

Yet another measure of the turbulence compensation quality is the ratio of compensation error energy to the original image energy. The error image's energy to the original image's energy ratio is given by:

$$\frac{\sum_{\vec{p}\in\Omega}\left[I_O(\vec{p}) - \hat{I}(\vec{p})\right]^2}{\sum_{\vec{p}\in\Omega} I_O^2(\vec{p})} \tag{4-6}$$

where $I_O(\vec{p})$ is the original image and $\hat{I}(\vec{p})$ is the image after inducing and compensating turbulence degradations, Ω is the image plane. For the given example, this *error-image's energy to the original-image's energy ratio is as low as 0.04*.

As described in Sect. 4.3 real motion extraction is based on assumptions made for the statistical behavior of turbulent motion. When those assumptions break, for example when winds speed up the turbulent motion and give it a strong regularity, turbulent motion will be regarded as real moving objects and will affect the visual output.

4.5.3. Turbulence Compensation - Qualitative Evaluation

In order to verify the method's real-life applicability, turbulent degraded sequences which were acquired by operational long range observation systems were processed by the system.

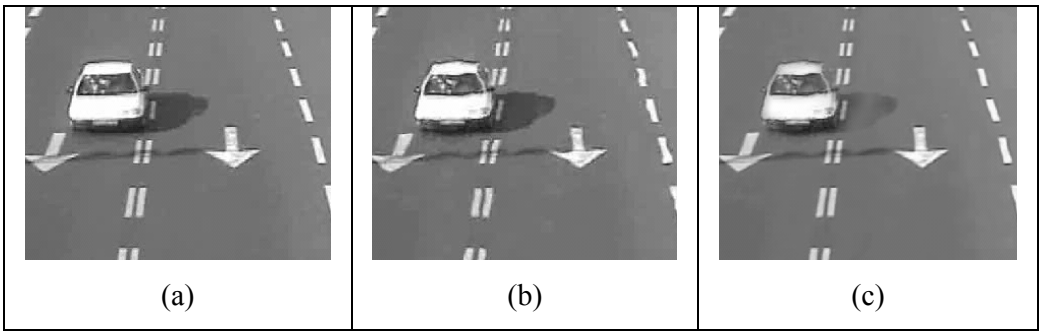


Figure 4-7 – Turbulence Simulation and Compensation. (a) Original Sequence with Real-Motion and no turbulent motion; (b) the same frame with turbulence-like distortions induced, using computer software; (c) the same frame after the turbulence compensation process.

Figure 4-8 and Figure 4-9 illustrate the results obtained with real-life video sequences. Figure 4-8(a) and Figure 4-9(a) are both frames extracted from real life turbulent degraded sequences (see [123]). Corresponding figures (b) show the stable scene estimation computed by element-wise temporal median over of 117 frames. Figures (c) display the Distance From Median Masks (DFM) and (d) display the Real Motion Separation Masks (RMSM). Comparing figures (c) and (d) in both Figure 4-8 and Figure 4-9, one can notice how the real motion extraction process is refined. Background areas which were tagged as real-motion by the DFM in (c) are removed from the RMSM in (d). The stable output of the non-turbulent background and unaffected vehicles with real moving objects are given in, Figure 4-8(e) and Figure 4-9(e).

4.6. Real-time Applicability

The number of operations, needed to complete a computational task is platform independent, while other measures, such as execution time, are resources and platform dependent. Therefore the real-time applicability is evaluated through the number of operations[1]. The analysis of the number of calculations in the following sections is given in terms of operations per pixel.

4.6.1. Stable Scene Estimation Computation

Generally, the rank filtering and, specifically, median filtering has high computational complexity since a sorting operation is carried out for each median

[1] A tool, which helps examining if the number of operations for an algorithm complies with real-time constrains on a specific machine, can be downloaded from the author's site [123]

computation. Exploiting the fact that the median computation is held over a temporal-wise sliding window allows utilizing a fast recursive method for median filtering. The fast median recursive computation method used is a variation of the method described in [129]. First it holds an initialization process, in which odd number of $\boldsymbol{N}$ frames comprising the temporal window are read and the temporal histograms are computed for each pixel.

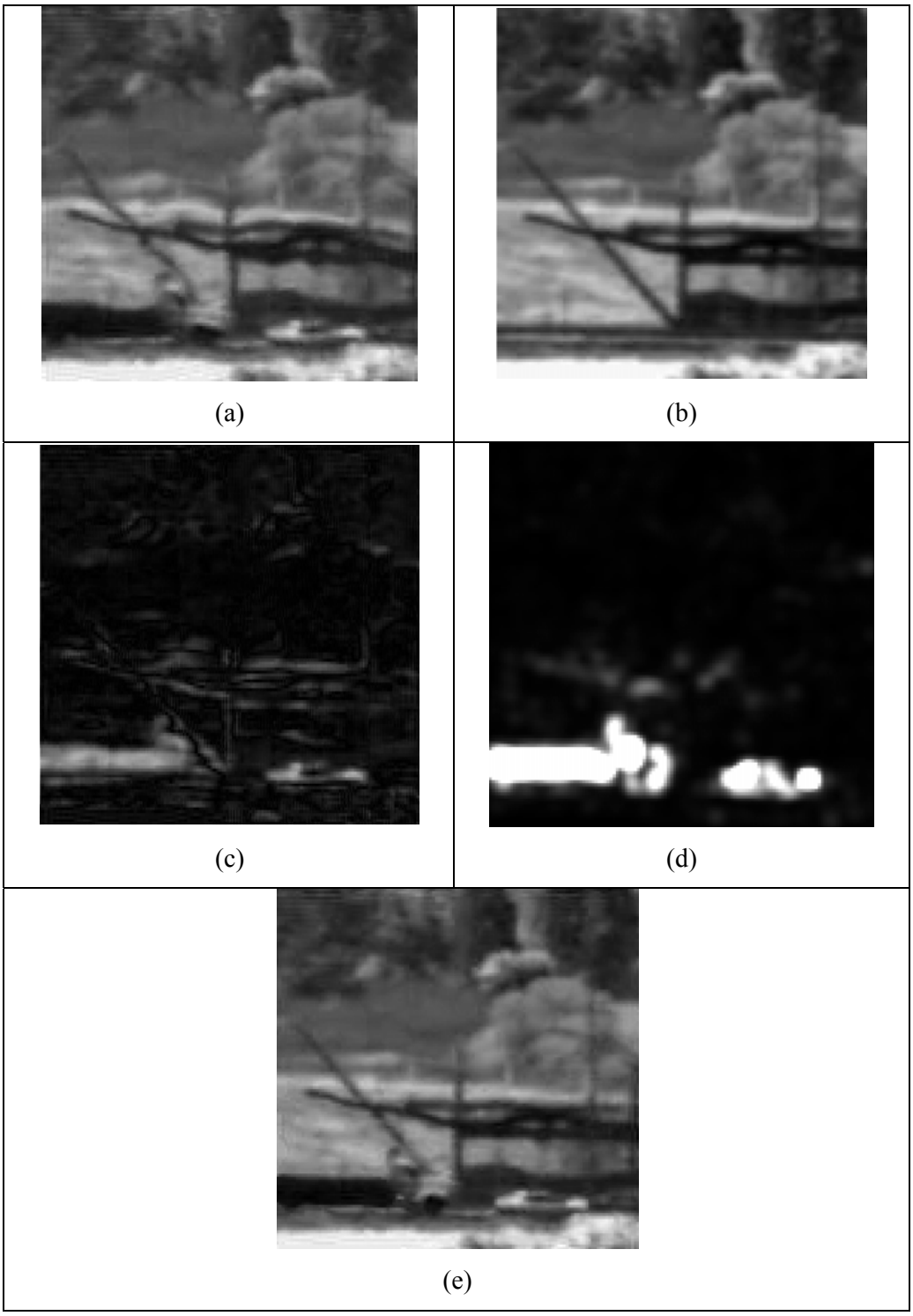


Figure 4-8 - Real-time Turbulence Compensation – Trucks Sequence. (a)-Real-life atmospheric turbulence degraded image; (b)- the stable scene estimation; (c)- DFM; (d) - RMSM; (e) - the output image

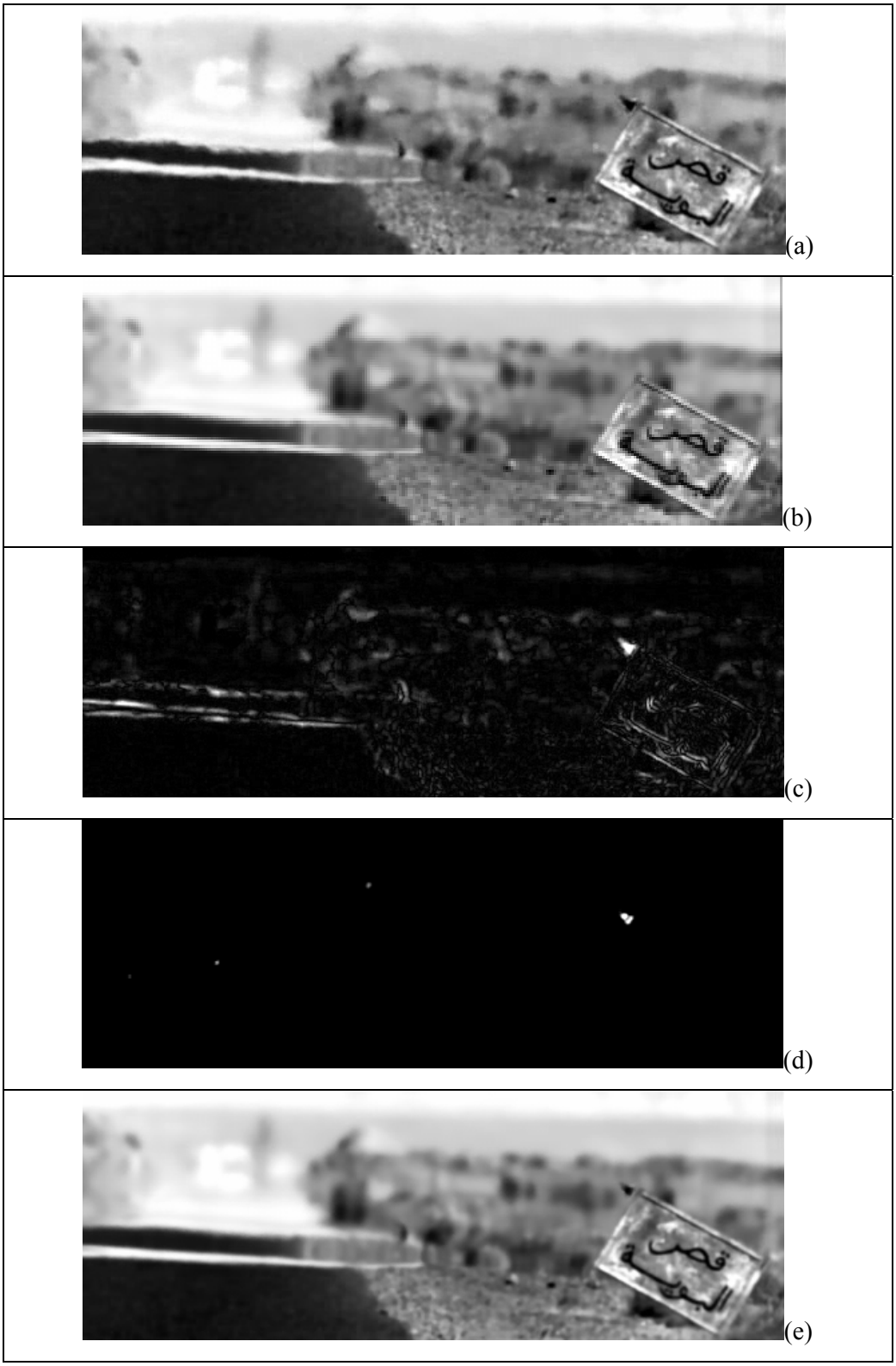


Figure 4-9 - Real-time Turbulence Compensation – Bird Sequence. (a)-real-life atmospheric turbulence degraded image; (b)-the stable scene estimation; (c)- DFM; (d)- RMSM; (e)- the output image.

Having completed the initialization process, for each pixel the following data is maintained: (a) The temporal window gray-level histogram; (b) The actual gray-level median value (***b***-parameter); (c) The number of pixels in the window having the same intensity level as the median (***c***-parameter); (d) The number of pixels in the temporal window, which have lower gray-level values than the median (***d***-parameter).

Upon new frame acquisition, for each pixel, these data are updated according to the new and departing values of the moving window. If the departing or new values equal to the median then the ***c***-parameter for that pixel is decreased or, respectively, increased by one; if the departing or new values have lower value than the current median then ***d*** -parameter is decreased or increased in the same manner. Upon completion of the update process, one can be determined if the conditions described by equations (4-7) and (4-8) are fulfilled.

$$d < \frac{N+1}{2} \tag{4-7}$$

$$[c+d] \geq \frac{N+1}{2} \tag{4-8}$$

If this is the case, then the pixel has preserved its median value (b). If this is not the case, for a given pixel, the median is calculated using the temporal histogram (a) and the pixel's corresponding ***b***-, ***c***- and ***d***- parameters are updated. Typically, the median values do not change significantly over time. Therefore the need for calculating the histogram for every pixel in every incoming frame will be notably smaller, thus decreasing the load derived due to the median computation.

Figure 4-10 contains three different frames (a), (c) and (e), taken in different times, from a real-life turbulent sequence (see [123]). The same location is marked with a cross on each of those images. Images (a), (c) and (e) are taken before, throughout and after the bird passes through the marked pixel. The corresponding temporal histograms, computed (over 117 frames) in a temporal window centered at the corresponding frame, of this location are given in figures (b), (d) and (f). The dashed line in those images represent the median's gray-level value. The median gray-level's standard deviation throughout the entire sequence was 3.05 gray-levels, which, for the human, is an insignificant difference. The median value itself rarely changes throughout the sequence. In this specific example, the median was recalculated for less than 1% of the pixels.

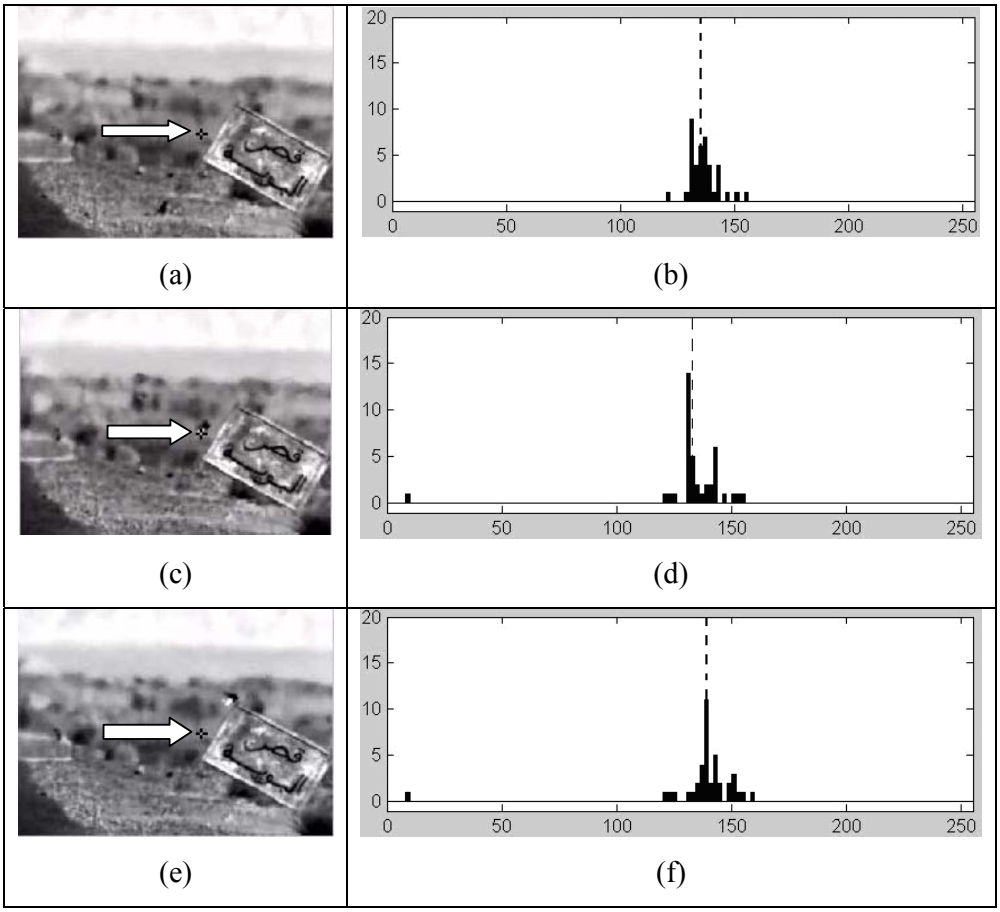


Figure 4-10 – Temporal Median. Figures (a), (c) and (e) are images taken from a turbulent degraded real-life sequence. Figure (b), (d) and (f) are the corresponding temporal histograms for the pixel marked by a cross and pointed by an arrow. The dashed line on the histogram represents the temporal median gray-level value.

4.6.2. Real Motion Extraction

The first real-motion discrimination phase consists of a simple subtraction of the median from the input frame and applying threshold on the result, therefore the number of operations per pixel is 2.

The second phase consists on optical flow computation. Optical Flow methods require the minimization of an energy functional (see Sect. 3.3). In order to solve these large sparse systems numerically, classical iterative methods such as the Gauss–Seidel algorithm are commonly used. While they are simple to implement, their convergence is not very fast, and often thousands of iterations may be necessary to get sufficiently close to the minimum of the energy functional. This is the reason why optical flow methods are too slow for time-critical applications.

As derived from the theory, the turbulence distortions are modeled by spatial, and temporal, random local shifts. This means that the motion of a pixel should be derived from its local neighborhood in the acquired and reference frames, rather than the entire image. While the general purpose of optical flow methods is computing the entire dense-map, it is suggested, for the sake of reducing the computation load, to apply the optical flow computations to only certain pixels rather than to the entire image. As pixels tagged as stationary by phase I are not processed by phase II, the number of pixels processed by this phase will be small enough to allow real-time applicability. If the number of pixels processed in this phase is too big so the computational complexity exceeds the processing machine capabilities, the estimation of phase I is used solely. The magnitude and angles computation and thresholding are regarded as part of the computations needed for optical flow.

4.6.3. Generation of the Output Frames

According to equation (4-4), the generation of the stable frames requires 8 operations per pixel.

4.6.4. Total Computational Complexity

Table 4-1 shows the number of operations per pixel for each task of the algorithm's tasks. As the number of pixels processed by phase II of the real motion extraction is content dependant, the maximum and minimum number of operations is indicated.

Table 4-1 – Number of Operations per Task

Task	Maximum	Minimum
Median	~770	7
Real Motion Extraction – Phase I	2	2
Real Motion Extraction – Phase II	~170	0
Generation of the Output Frames	8	8
Total	~950	17

The acquisition system, acquires interlaced 4CIF format images of 704x576 pixels [130,131]. The method disclosed here eliminates the need for computing the entire dense-map, applying a hierarchical motion segmentation technique. If the number of pixels processed on the second stage of the real-motion extraction is less than 10% of the total number of pixels in the frame then the performance is equivalent

to processing 25 frames per second on a standard 3.0-GHz PC. Further improvement to processing time can be achieved by utilizing the interlaced property of the incoming video. This is achieved by processing the odd and even fields individually on multi-processor/multi-core architectures.

4.7. Summary

A real-time video stabilization scheme which preserves real motion in the scene was presented. For preserving real motion in the scene, moving objects are located and the compensation for the distortion of the turbulence is applied only to the static areas of images. To this goal, for each pixel in the incoming frame it is decided whether it is of a moving or a stationary object. This is done by a hierarchical two-stage decision making mechanism. In order to evaluate the algorithm's performance envelope, computer simulation software was written. The software induces turbulence-like distortions on any input sequence. Finally, the real-time applicability of the algorithm was discussed and the condition under which the algorithm complies with real-time constrains were formulated. While the algorithm can be utilized in various applications, the discussion in this chapter revolved around motion presented by turbulence or camera's random ego motions. The utilization of the described method in a different class of applications is described in Chapter 7.

5. Video Resolution Enhancement

A frequent distortion, in videos, is image instability in the form of chaotic global and local displacements. Those degraded videos contain tremendous redundancy that potentially can be used for image resolution enhancement, through elastic registration with sub-pixel accuracy, of segments of video frames that represent stable scenes. In this chapter two novel practical super-resolution schemes for monochrome and color turbulent degraded videos and their utilization in real-time are presented. More than that, the potentials and limitations of such resolution enhancement schemes are studied by means of the achievable resolution enhancement possible, under given imaging setup parameters, such as the camera fill factor, the intensity of pixel displacements and of the number of image frames. This study results in useful insights for the design of such systems.

5.1. Super-resolution in Turbulent Videos Principals

In turbulence-corrupted videos, consequent frames of a stable scene differ only due to small atmospheric-turbulence-induced movements between images. As a result, the image sampling grid defined by the video camera sensor may be considered to be chaotically moving over a stationary image scene. This phenomenon allows the generation of images with larger number of samples than those provided by the camera if image frames are combined with appropriate re-sampling [49,50,127,132,133,134].

Generally, such a SR process consists of two main stages [26,33,37 ,42,127,133,134]. The first is determination, with sub-pixel accuracy, of pixel movements. The second is combination of data observed in several frames in order to generate a single combined image with higher spatial resolution. A flow diagram of this stage of processing is shown in Figure 5-1.

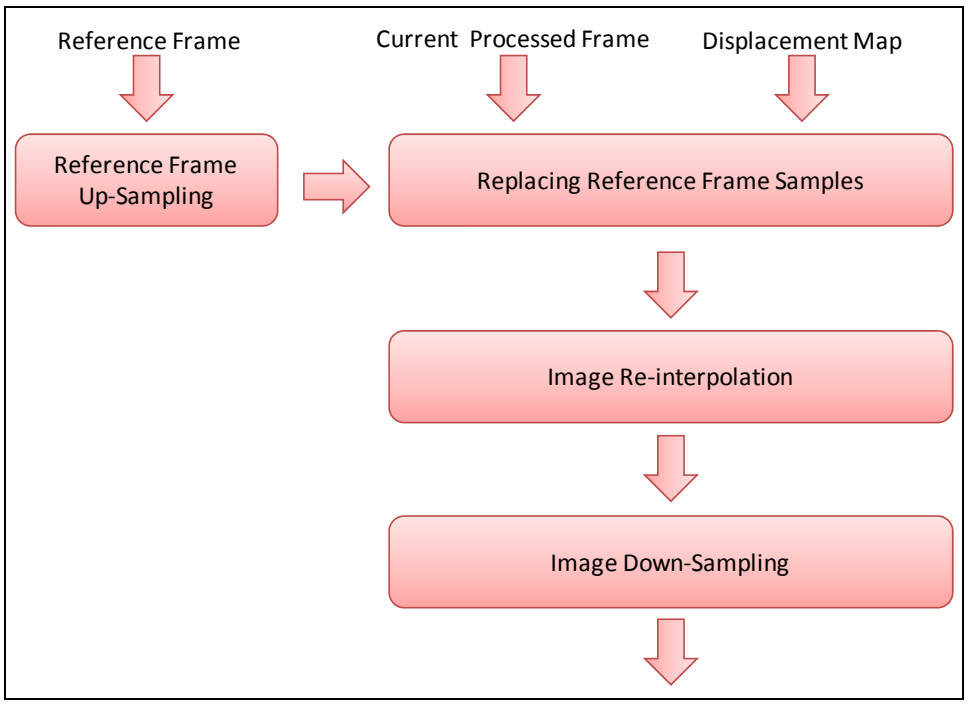


Figure 5-1 - Flow diagram of the process of generation of stabilized frames with SR

For each current frame of the turbulent video, inputs of the process are: a corresponding reference frame, obtained as a pixel-wise temporal median over a time window centered on the current frame (see Sect. 4.2), and the current frame displacement map. The latter serves for placing pixels of the current frame, according to their positions determined by the displacement map, into the reference frame, which is correspondingly up-sampled to match the sub-pixel accuracy of the displacement map. For up-sampling, different image interpolation methods can be used. Among them, discrete sinc-interpolation is the most appropriate as the one with the least interpolation error and may also be computed efficiently [118]. As a result, output stabilized and enhanced, in its resolution, frame is accumulated. In this accumulation process it may happen that several pixels of different frames are to be placed in the same location in the output enhanced frame. In order to make best use of all of them, the median of those pixels is computed in order to avoid the influence of outliers that may appear due to possible errors in the displacement map. The data accumulation process is illustrated in Figure 5-2, where figures (a), (b), (c) and (d) represent the accumulated data after the 1, 25, 100 and 200 frames. For the sake of clarity, pixels, which were not substituted in this process, are represented in black pixels.

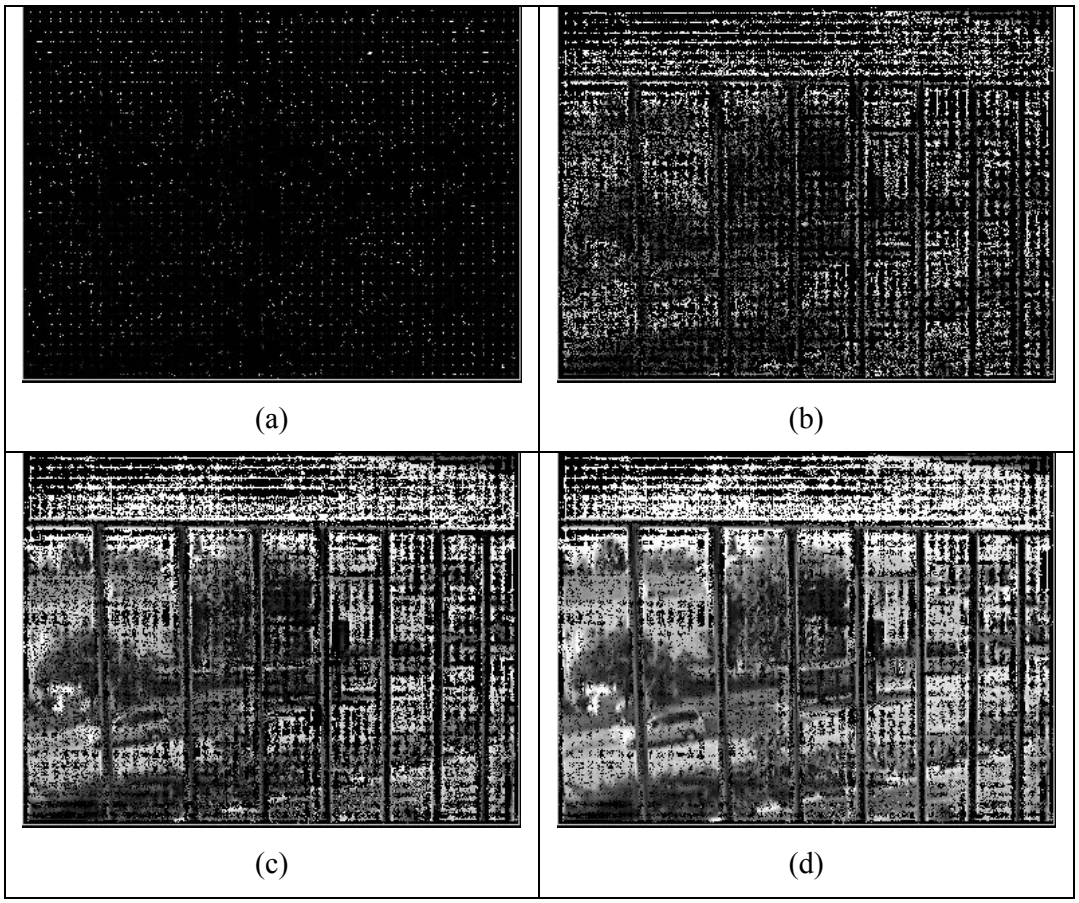


Figure 5-2 – Super-resolution data accumulation process. Figure (a) is the data accumulated after the first frame; figure (b) represents the accumulated data after 25 frames and figures (c) and (d) are the representations of the data accumulated over 100 and 200 frames, correspondingly. Pixels, which were not substituted, in this process, are represented in black pixels.

After all available input frames are used in this way, the enhanced and up-sampled output frame contains, in positions where there were substitutions from input frames, accumulated pixels of the input frames and, in positions where there were no substitutions, interpolated pixels of the reference frame. Substituted pixels introduce to the output frame high frequencies outside the base-band defined by the original sampling rate of the input frames. Those frequencies were lost in the input frames due to the sampling aliasing effects. Interpolated pixels that were not substituted do not contain frequencies outside the base-band. In order to finalize the processing and take full advantage of the SR provided by the substituted pixels, the iterative re-interpolation algorithm, depicted in Figure 5-3 was used [135, 136].

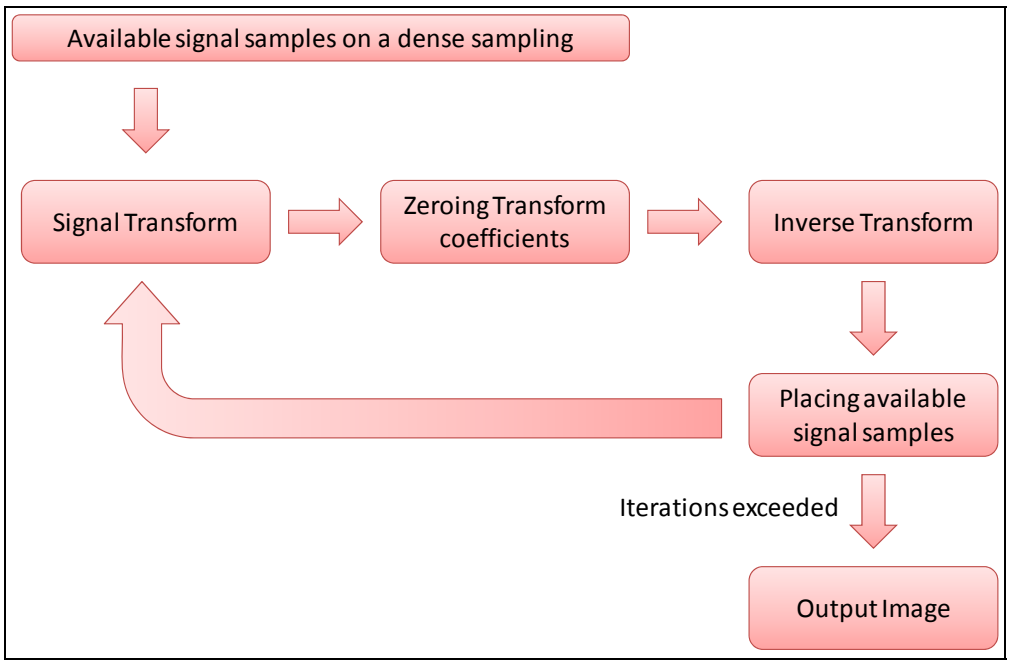


Figure 5-3 - Flow diagram of the iterative signal recovery procedure

This algorithm assumes that all substituted pixels accumulated, as described above, are stored in an auxiliary replacement map containing pixels values and coordinates. At each iteration of the algorithm, a transform of the image obtained at the previous iteration is computed and then zeroed in all its components outside of the selected enhanced bandwidth, say, double of the original one. After this, the inverse transform is performed on the modified spectrum and corresponding pixels in the resulting image are replaced with pixels from the replacement map thus producing an image for the next iteration. In this process, the energy of the zeroed outside spectrum components can be used as an indicator when the iterations can be stopped. The iterative signal recovery process is illustrated in Figure 5-3, where the output after 1, 5, 15 and 75 iterations is depicted in figures (a), (b), (c) and (d) respectively. The considerations for which transform to use are discussed in Ch. 6.

Once iterations are stopped, the output-stabilized and resolution-enhanced image obtained in the previous step is sub-sampled to the sampling rate determined by selected enhanced bandwidth and then subjected to additional processing aimed at camera aperture correction and, if necessary, denoising.

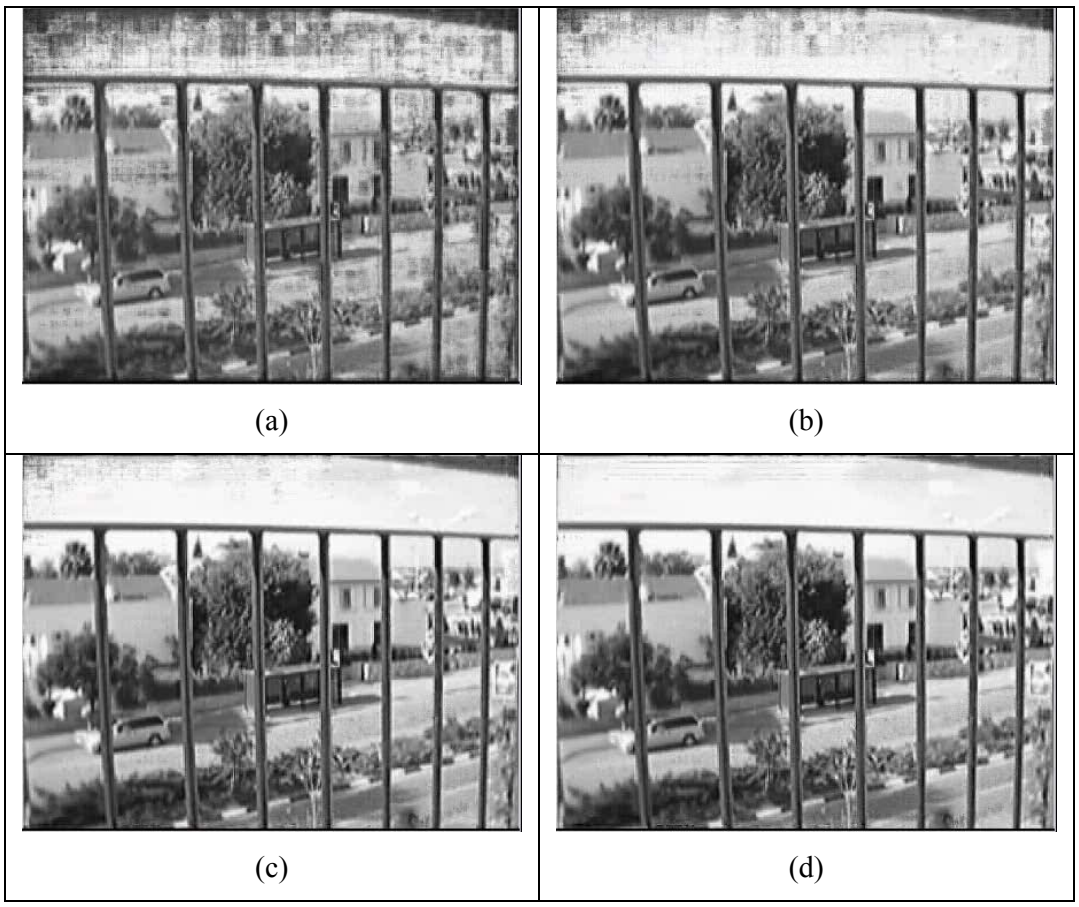


Figure 5-4 – The visual output of the iterative signal recovery process after 1 (figure (a)), 5 (b), 15 (c) and 75 (d) iterations.

5.2. Objective Criterion of Image Sharpness

5.2.1. Preface

Image evaluation, quality measurements and characterization are fundamental components in all image processing applications and techniques. The emergence of new technologies for displays panels, cameras and mobile devices and the growth in the number of manufacturers have emphasized the need for comparison and evaluation techniques, especially in the evolution of image restoration and compression algorithms. There are basically two classes of objective quality or distortion assessment approaches. The first are mathematically defined measures of error between the evaluated image and its ideal prototype. Typical examples of such measures are MSE, peak signal to noise ratio (PSNR), root mean squared error (RMSE), mean absolute error (MAE), and signal-to-noise ratio (SNR) [137]. The second class of measurement methods considers human visual system (HVS)

characteristics in an attempt to incorporate perceptual quality measures [138]. The required ideal image or a model, as a reference that the evaluated image can be compared to, in many cases, is not available.

Visual quality assessment through image spectra is widely used and plays a significant role in image assessment applications [138,139]. Additionally, image acquisition systems' quality is commonly determined by the optics or detector cutoff frequencies and bandwidth. However, none of the methods suggest a quantitative scalar evaluation for the effective image bandwidth. One can not compare two acquisition devices, scanners for example, and have a scalar criterion.

Image quality metrics that do not require a reference can be found in the literature [140]. The quality assessment is based, in those methods, on the evaluation of image edge sharpness. These metrics are sensitive to the choice of the edge detector, as well as to the presence of noise. To be edge detector invariant, a method which estimates image quality through the image blur effect was suggested [141]. The method compares the variations between neighboring pixels of the image before and after low-pass filtering. High variation between the original and the blurred image means that the original image was sharp whereas a slight variation between the original and the blurred image means that the original image was already blurred. The perceptual blur measure (PBM) ranges from 0 to 1 has shown to very well correlate with subjective evaluation of image sharpness degradation with 0 corresponding to the lowest sharpness degradation and 1 to the highest degradation [141].

This Sect. presents a new image quality analysis method, derived from the PBM-method. The method suggests a new image quality attribute, *image effective bandwidth (IEBW)*, which connects the image's energy distribution in the frequency domain with perceptual image quality.

5.2.2. Image Effective Bandwidth

In order to obtain numerical data that one can use to associate image bandwidth with PBM, the set of test pseudo-random images with uniform spectrum in certain fraction of the base-band, which is described in Sect. 3.4, is suggested. Five test sets, each consisted of 256 images with different bandwidths, were generated. Hence, for a given bandwidth, α, there were 5 different test images, one from each set. Those five images are referred to as bandwidth-group. The PBM was computed for each of the images. The PBM average of each bandwidth-group is depicted in Figure 5-5. In our

tests the average PBM standard deviation for the five testing sets is 0.0006. For a larger number of test sets the standard deviation practically converges to zero. As one can see, the function described in Figure 5-5 is a monotonic one, which implies a direct association of effective bandwidth with a specific PBM value. The inverse relation can be used for estimation of image's effective bandwidth from its PBM. A complete Matlab software package that computes image's PBM and corresponding effective bandwidth can be downloaded from [142].

The PBM computation process is based on local features of the image. This means that the size of the image has no effect on the PBM. Figure 5-6 shows the PBM computation for two test sets. The first set, is one of the five sets used to generate Figure 5-5 and it consists of 256x256 images with different 256 bandwidths. The second set consists of 512x512 images with 512 different bandwidths. As the effective bandwidth is normalized to the image base-band, both data sets can be put on the same graph, which is presented in Figure 5-6, where, for better presentation, the $PBM^{0.7}$ as a function of the effective bandwidth is shown for the 256x256 (solid) and 512x512 (dashed) test sets. Since the difference is barely noticeable, a fragment of the graph is magnified. The average difference, in our tests, was 0.0007. This implies that the IEBW measure is universal and can be used to compare effective bandwidths of images with different sizes.

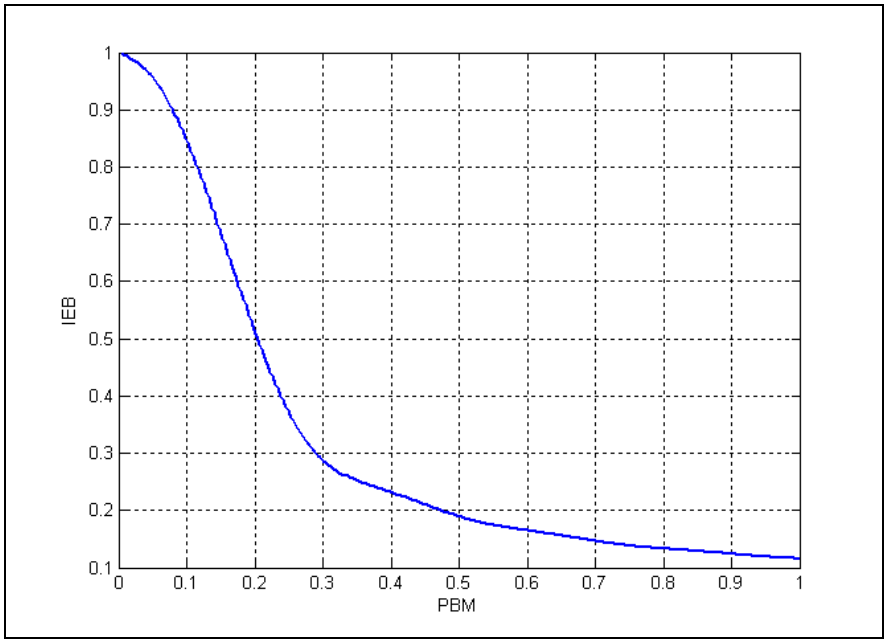


Figure 5-5 – The Average of the PBM computed for five test sets with known IEB

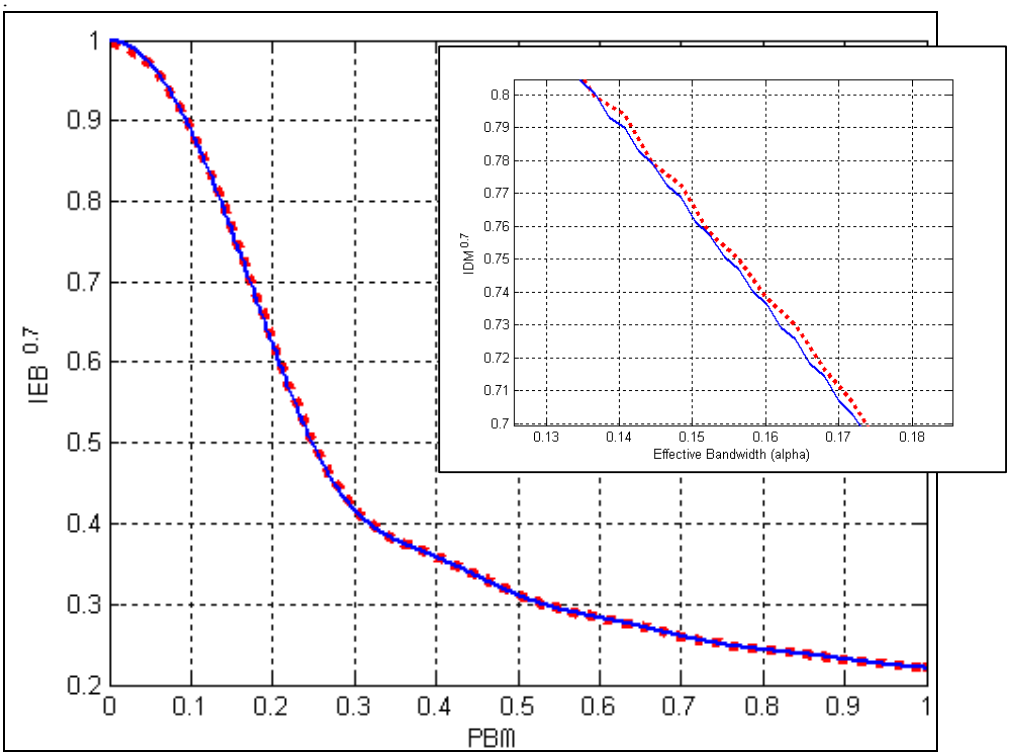


Figure 5-6 – The PBM computed for two test sets. The PBM of 256x256 pixels images is shown in solid line, while the PBM as a function of the effective bandwidth of the 512x512 test set is presented in dashed line.

5.2.3. Results

For validating the effective bandwidth measure, the four real-life images, presented in Figure 5-7, were used. The PBM and the effective bandwidth, for each image, are given in Table 5-1, which suggests that the houses image have the widest effective bandwidth, the panorama image has an intermediate bandwidth followed by the Lenna and the Peppers image which has the narrowest bandwidth of the four. This is supported by Figure 5-8, where the energy of the row-wise DCT coefficients normalized to the image total energy for the four real-life test images is depicted. In all graph the arrows point the corresponding image's IEBW. As one can see higher IEBW measures do correspond to larger portion of the image energy in its higher frequencies. This is presented quantitatively in

Table 5-2, where the total energy portions sited in the higher half of the DCT base-band as well as the portion of the energy from the images' IEBW onwards are given. Those results support the notion that the PBM measure suggested in [141] is suitable for measuring the image effective bandwidths.

Figure 5-7 – 4 Real-life images used in the validation process: Houses, Panorama, Lenna and Peppers

5.3. Super-resolution in *Monochrome* Videos

Figure 5-9 illustrates the feasibility of the method. Figure (a) is a frame extracted from turbulent degraded real-life sequence, while figure (b) is its super-resolved stable one. Figures (c) and (d) are corresponding magnified fragments from figures (a) and (b). The fragments are marked with black boxes on (a). In both, figures (c) and (d), the original fragments are shown on the left-hand side, while the super-resolved fragments are shown on the right-hand side.

Atmospheric turbulence also affects thermal range videos. Figure 5-10 demonstrates application of the method to an intermediate infrared wavelengths (3-8 μm), turbulent video sequence. Figure 5-10 (a) is a frame extracted from turbulent degraded thermal video, while figure (b) is the corresponding super-resolved frame generated from the thermal sequence. The marked fragments of figure (a) are presented in figures (c) and (d), in which fragments with initial resolution are given on the left-hand side, while the super-resolved fragments are given on the right-hand side.

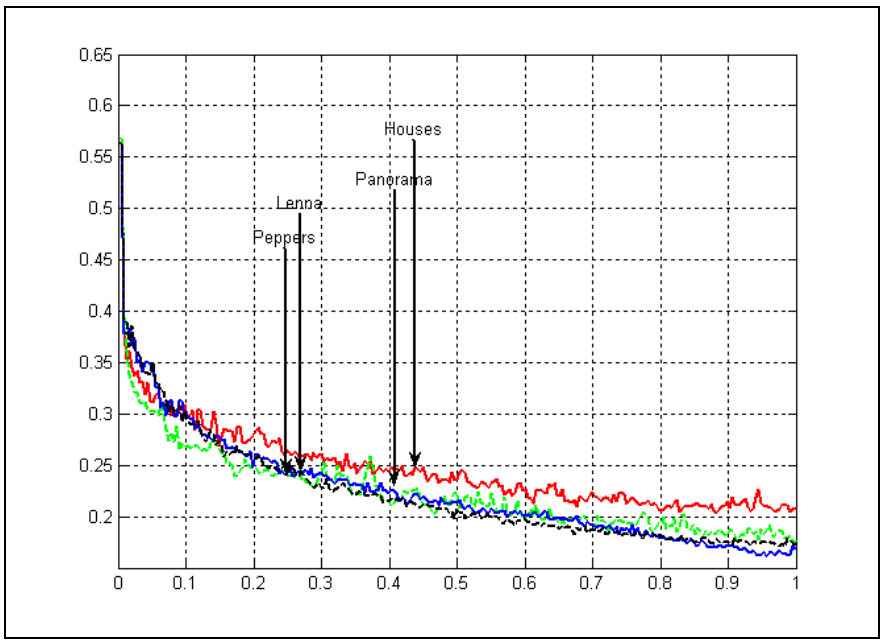


Figure 5-8 – energy of the row-wise DCT coefficients normalized to the image total energy for the four real-life test images. the arrows point the corresponding image's IEB.

Table 5-1 – Real-life images PBM and effective bandwidth measures

Image	Houses	Panorama	Lenna	Peppers
PBM	0.21	0.23	0.33	0.37
Effective Bandwidth	0.44	0.41	0.27	0.25

Table 5-2 - Energy of higher frequencies DCT coefficients as a fraction of the total energy of images and their effective bandwidth

Image	Houses	Panorama	Lenna	Peppers
Effective Bandwidth	0.44	0.41	0.27	0.25
Normalized energy of higher half of the DCT coefficients	0.0183	0.0102	0.0038	0.0036
Normalized energy of the higher DCT coefficients beginning from IEB	0.0231	0.0142	0.0112	0.0132

For quantitative evaluation of the resolution improvement we used a method for numerical evaluation of image effective bandwidth (IEBW) suggested, by the author in [143], and thoroughly explained in Sect. 5.2. The IEBW factors for the images shown in Figure 5-9 and Figure 5-10 are given in Table 5-3, which suggests that the

super-resolved images do have higher effective bandwidths, hence more information in higher frequencies.

Table 5-3 – Quantitative evaluation of the super-resolved images through Image Effective Bandwidth (IEBW)

	Original	**Super-Resolved**
Visual Range video - Figure 5-9		
Entire Original Image (Figure 5-9(a)) Vs. Entire Super-Resolved Image (Figure 5-9(b))	0.1645	0.2366
Fragment (c)	0.1044	0.174
Fragment (d)	0.0983	0.1868
Thermal-range video - Figure 5-10		
Entire Original Image (Figure 5-10 (a)) Vs. Entire Super-Resolved Image (Figure 5-10 (b))	0.1343	0.188
Fragment (c)	0.1067	0.1492
Fragment (d)	0.1542	0.1998

5.4. Super-resolution in *Color* Videos

In this section, the theory laid for the monochrome case is extended to turbulent *color* videos and it is shown that, while SR in color images is somehow limited, due to the color video acquisition mechanism, compared to gray-scale images, image resolution enhancement is achievable as well.

A color image can be represented by combining three separate monochromatic images. Ideally, each image pixel contains three data measurements; one for each of the three color bands, R, G and B. In practice, commonly used digital camera with a single charge coupled device (CCD) array provides only one color measurement (red, green, or blue) per pixel. The detector array, in such cameras, is a grid of CCDs, each made sensitive to one color by placing a color filter array (CFA) in front of the CCD. The Bayer pattern, shown on Figure 5-11, is a very common example of such a color filter arrangement. The values of missing color bands at every pixel are synthesized using some form of interpolation from neighboring pixel values. This process is referred to as color demosaicing.

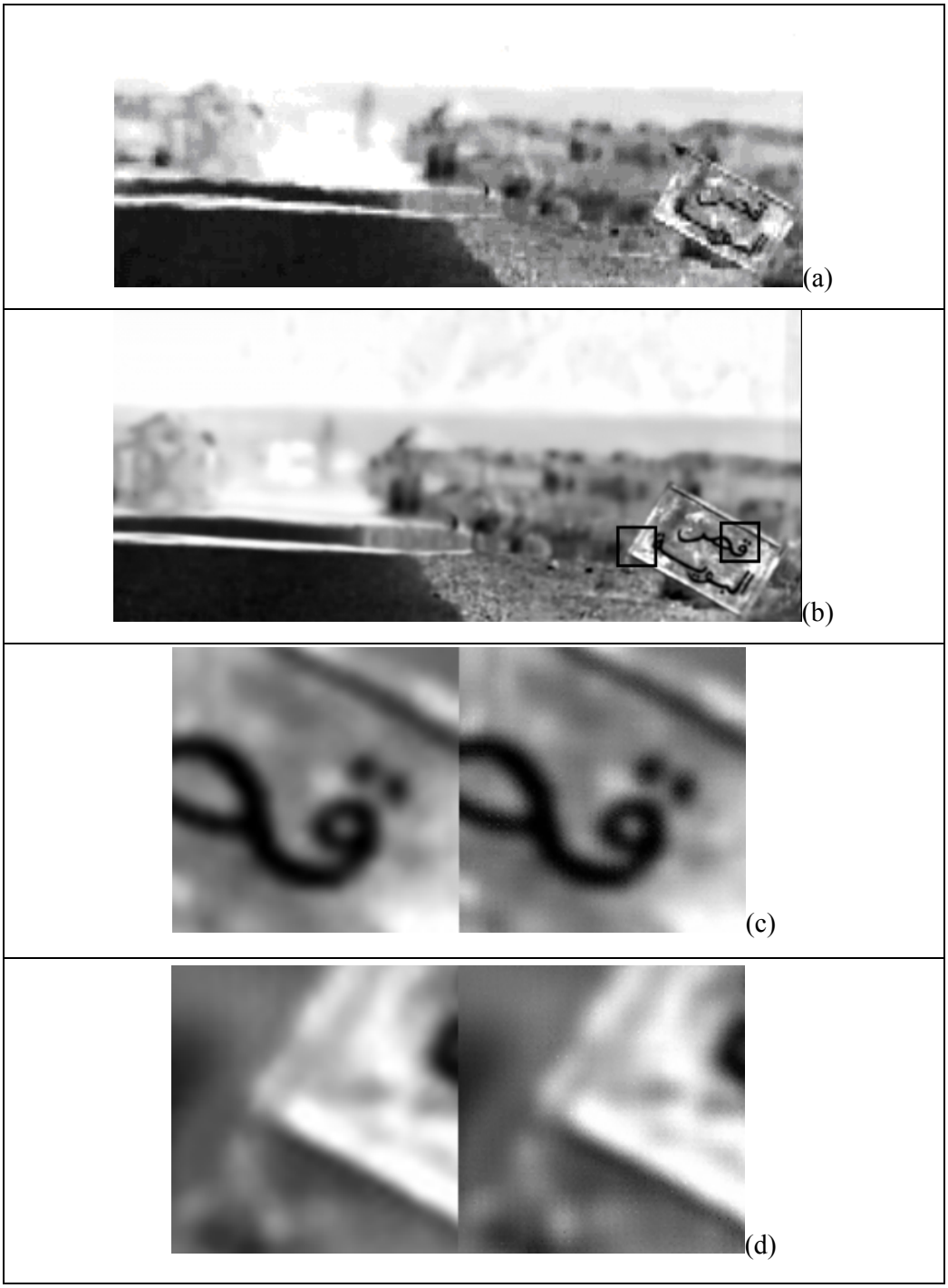


Figure 5-9 - SR through Turbulent Motion – Visual-range Sequence. (a) shows a raw video frame, (b) shows a super-resolved frame generated from a visual range turbulent degraded real-life video. (c) and (d) are the magnified fragments marked on (b) – the left-hand side shows the fragment with simple interpolation of the initial resolution and the right hand side shows the fragment with super–resolution.

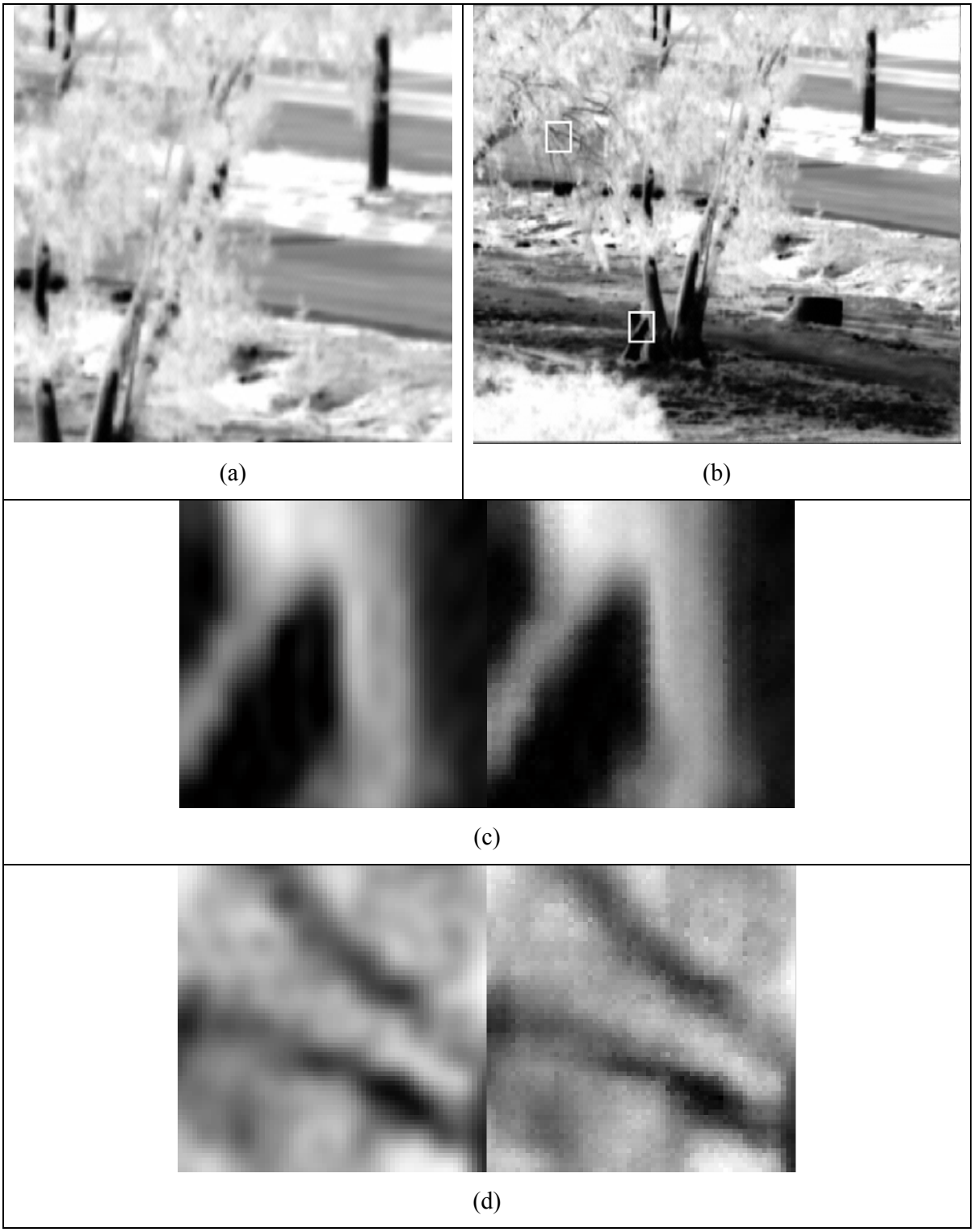


Figure 5-10 - SR through turbulent motion. (b) is the super-resolved frame generated from a thermal range turbulent degraded real-life video, which one of its frame is presented in figure (a). Figure (c) and (d) are the magnified fragments marked on (b) – the left-hand side shows the fragment with simple interpolation of the initial resolution and the right hand side shows the fragment with super –resolution.

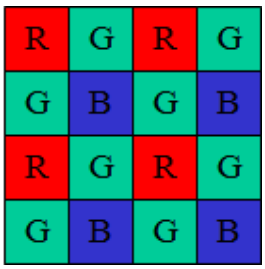


Figure 5-11 – Bayer Color Filter Array (CFA)

Linear demosaicing methods can be represented by three sets of weights for neighborhoods with R,G or B pixel in the center. For (x,y)-th pixel it is given by:

$$I_{(x,y)}^{R,G,B} = \sum_{\Delta_1,\Delta_2 \in \Omega} w_{(\Delta_1,\Delta_2)}^{R,G,B} C_{(x+\Delta_1,y+\Delta_2)}^{R,G,B} \tag{5-1}$$

where $I_{(x,y)}^{R,G,B}$ are either the R,G or B interpolated values, $C_{(x+\Delta_1,y+\Delta_2)}^{R,G,B}$ are the R,G and B sampled output values of CCD cells covered by the Bayer filter, Ω is a neighborhood centered at (x,y) and $w_{(\Delta_1,\Delta_2)}^{R,G,B}$ are the corresponding interpolation R,G and B weights normalized so as

$$\sum_{\Delta_1,\Delta_2 \in \Omega} w_{(\Delta_1,\Delta_2)}^{R,G,B} = 1 \tag{5-2}$$

Applying the Fourier transform on (5-1) gives the modulation transfer function (MTF) of the interpolation procedure:

$$H_{(r,s)}^{R,G,B} = H_{(r,s)} \sum_{\Delta_1,\Delta_2} w_{(\Delta_1,\Delta_2)}^{R,G,B} \cdot e^{-2\pi i (r\cdot\Delta_1 + s\cdot\Delta_2)} \tag{5-3}$$

where r and s are the image spatial frequencies (see Sect. 2.1) and $H_{(r,s)}$ is the MTF of monochrome 2D rectangular pixel [93,144]. Considering (5-2), one can derive that:

$$\left| \sum_{\Delta_1,\Delta_2 \in \Omega} w_{(\Delta_1+\Delta_2)}^{R,G,B} \cdot e^{-2\pi (r\cdot\Delta_1 + s\cdot\Delta_2)} \right| \le 1 \tag{5-4}$$

which implies that (5-3) is bounded by $H_{(r,s)}$, hence:

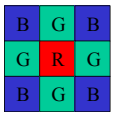
$$\left\{ \left| H_{r,s}^{R,G,B} \right| \le \left| H_{(r,s)} \right|; \forall (r,s) \right\} \tag{5-5}$$

Equations (5-3) and (5-5) show that the color MTF for R, G or B pixels is a weighted average of monochrome MTF's, hence the camera fill factor, which determines the MTF, has the same impact on the achievable SR results in color images as in monochrome ones. Additionally, (5-5) suggests that the achievable results for color images are bounded by the gray-level ones.

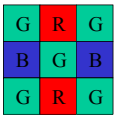
In commonly used color video cameras, JPEG image compression standard is used in which color images are first transformed from raw RGB-representation to the YCbCr representation [145]. In this representation, most of the energy is concentrated in the luminance (Y) component, with little energy in the chrominance (Cb, Cr) components [33]. Therefore, for processing color turbulent videos one can apply the above described monochrome SR algorithm to only the Y component, and use simpler processing of the Cb and Cr components. In our proposed method, the Cb and Cr components are computed using a pixel-wise temporal median filtering, which was shown to be a good estimation of the stable scene (see Sect. 4.2). For JPEG images, the Y component is given by [146]:

$$Y = 0.299\,R + 0.587\,G + 0.114\,B \tag{5-6}$$

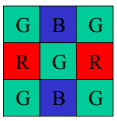
The Y component's MTF depends on the interpolation method. In JPEG encoding standard, the bilinear interpolation is used and each color channel is interpolated independently. Substitution of (5-6) into (5-3) yields the Y components' MTFs for that the cases when the central pixel is red, green and blue, respectively:



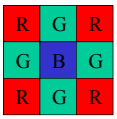
$$Y^{R}_{(r,s)} = H_{(r,s)} \cdot \left[\begin{array}{l} 0.299 + 0.2935\left[\cos\left(2\pi \frac{r}{N}\right) + \cos\left(2\pi \frac{s}{M}\right) \right] + \\ 0.114 \cos\left(2\pi \frac{r}{N}\right) \cdot \cos\left(2\pi \frac{s}{M}\right) \end{array} \right] \tag{5-7}$$



$$Y^{G}_{(r,s)} = H_{(r,s)} \cdot \left[0.299 \cos\left(2\pi \frac{r}{N}\right) + 0.587 + \frac{0.114}{2} \cos\left(2\pi \frac{s}{M}\right) \right] \tag{5-8}$$



$$Y^{G}_{(r,s)} = H_{(r,s)} \cdot \left[0.299 \cos\left(2\pi \frac{s}{M}\right) + 0.587 + 0.057 \cos\left(2\pi \frac{r}{N}\right) \right] \tag{5-9}$$



$$Y^{B}_{(r,s)} = H_{(r,s)} \cdot \left[\begin{array}{l} 0.299 \cos\left(2\pi \frac{r}{N}\right) \cdot \cos\left(2\pi \frac{s}{M}\right) + \\ 0.2935\left[\cos\left(2\pi \frac{r}{N}\right) + \cos\left(2\pi \frac{s}{M}\right) \right] + 0.114 \end{array} \right] \tag{5-10}$$

where for each equation the corresponding Bayer scheme setup is illustrated and N and M are the image dimensions for each axis.

One important conclusion derived from (5-7) through (5-10) is that the MTF of the interpolation process is shift variant. The different Y MTFs are depicted in Figure 5-12, where the Y MTF for red, green and blue are given in solid red, dotted green or cyan and dash-dot blue plots, which correspond to Eq. (5-7), Eqs. (5-8) and (5-9) and Eq. (5-10). As one can see, for neighborhoods centered in blue or red pixels, some frequencies of the Y component are completely lost (marked with arrows on figures (b) (c) and (d)). In the case at hand, the acquisition is held under turbulent conditions; therefore the same information is likely to be acquired by adjacent, but different, pixels in subsequent frames. This means that frequencies totally suppressed in a certain pixel, are likely to be recovered when the same information is acquired by neighboring pixel with different color filter (red, green or blue), in a subsequent frame. A similar idea was suggested in [71].

The entire processing was experimentally tested on real-life turbulent degraded color videos. The experimental setup is illustrated in Figure 5-13. Colored tiles are laid on the bottom of a water tub, filled with water to the height of 30 cm, while the turbulent motion of the water is caused by the filling water. The light propagates through the water turbulent medium and is acquired by the single-CCD camera, which has Bayer filter on its sensor. The results are shown for two different sequences: The first, referred to as the 'duck' sequence, is 100x160 pixels in size with average inter-frame turbulent motion magnitude of 1.9 and standard deviation of 1.19 pixels. The second sequence, the 'star' sequence, is 200x200 pixels with average turbulent inter-frame motion magnitude of 2.5 and standard deviation of 1.3 pixels.

Figure 5-14(a) presents a fragment of a turbulent deformed frame from the 'duck' sequence. Figure 5-14(b) is the estimation of the stable scene, computed over 300 frames of the original sequence and interpolated to twice of its original size. Figure (c) is the resolution enhanced image after applying aperture correction by unsharp masking. As visually resolution improvement can be appreciated only on a high quality display, image in Figure 5-14(d) presents the difference between images of figures (b) and (c) that clearly shows edges are enhanced. For allowing better evaluation, figures (d) through (h) represent corresponding fragments extracted from figures (a) – (d) respectively.

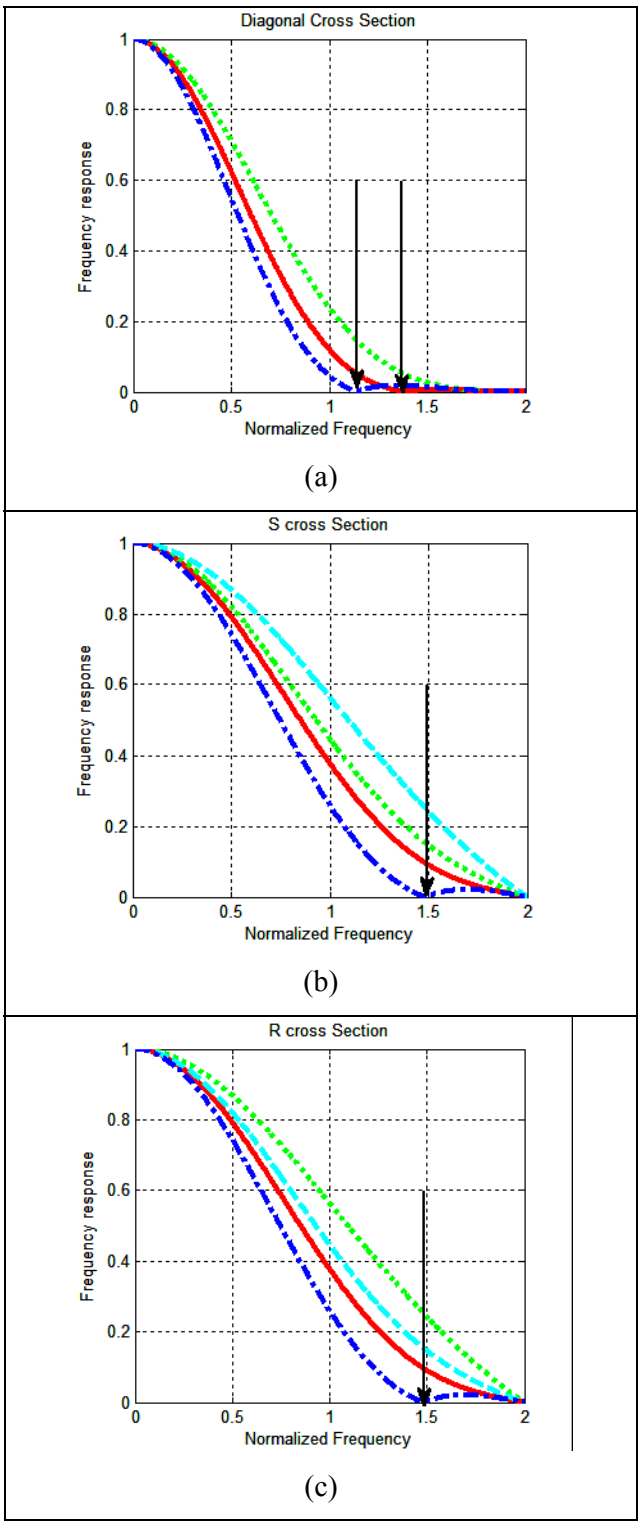


Figure 5-12 - Sensor frequency response for fill factor of 1 – 2D and cross Sect. views. Figures (a), (b) and (c) are the cross sections of the Y components frequency responses along the diagonal, R and S axes respectively; the green and cyan, red and blue lines represent the Y component frequency response for green, red and blue pixels in the Bayer scheme. The arrows mark the frequency where for the blue or red channel all luminance information is lost.

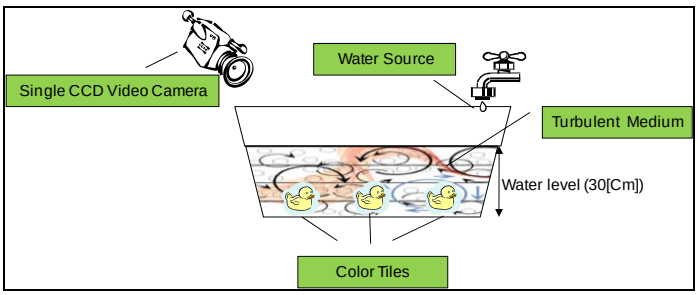


Figure 5-13 – Acquisitioning Setup

Figure 5-15 illustrates the efficiency of the SR process through images' spectra. Figure 5-15(a) presents the estimation of the stable scene of a turbulent degraded frame extracted from the 'star' video, interpolated to twice of its original size. The corresponding super-resolved image, after aperture correction, is depicted in Figure 5-15(b). Figures (c) and (d) show corresponding image spectra intensities. It is evident that information in higher frequencies outside the base band of original images is reconstructed in the process.

For quantitative evaluation of the IEBW measure was used (See Sect. 5.2). The IEBW factors for the images shown in Figure 5-14 and Figure 5-15 are given in Table 5-4, which suggests, similarly to the monochrome case, that the super-resolved images do have higher effective bandwidths.

In conclusion, one can state that the presented results confirm that color sequences subjected to turbulence and acquired by commonly used camera with single CCD array can be considerably enhanced in their resolution using the SR process.

Table 5-4 – The effective bandwidth measure for Figure 5-14 and Figure 5-15 for the interpolated and super-resolved images, both are twice in size than the original image

	Figure 5-14	Figure 5-15
Interpolated Image	0.2050	0.1459
Super-Resolved Image	0.2074	0.1814

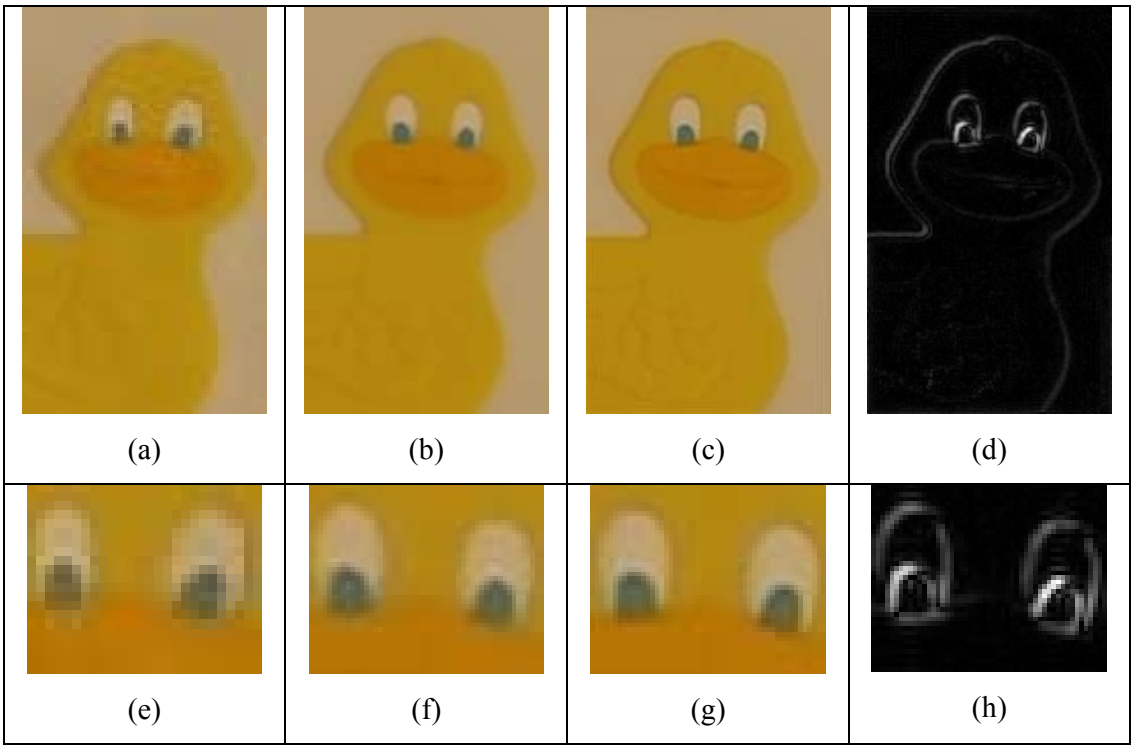


Figure 5-14 – The SR process. Figure a) is the original turbulence degraded color image, (b) is the reference frame computed in the first processing stage, interpolated to twice of its original size. Figure (c) is the super-resolved image Figure e) is the absolute difference of (b) and (c). Figures (e)-(h) are a magnified matching fragment from figures a) through d) respectively.

5.5. Potential and Limitation

In the evaluation of the results obtained in the previous sections, for real-life video SR, one should take into account several parameters such as variance of turbulence local displacements, fill factor of video camera, the number of frames involved in the process and the number of re-interpolation iterations. This section presents results of investigation, by means of computer simulation, into how those parameters affect the SR results.

5.5.1. Computer Model

In order to understand the potentials of SR from distorted video, the computer simulation software, presented in 4.5, was extended so it could examine the influence of the different parameters on the maximum achievable performance. The following parameters were investigated: Camera fill factor, turbulence intensity, the number of input frames, the number of iterations in the process of re-interpolation. Those are

discussed, correspondingly, in sections 5.5.1.1-5.5.1.4. The generation of the output synthetic frame is described in Sect. 5.5.1.5.

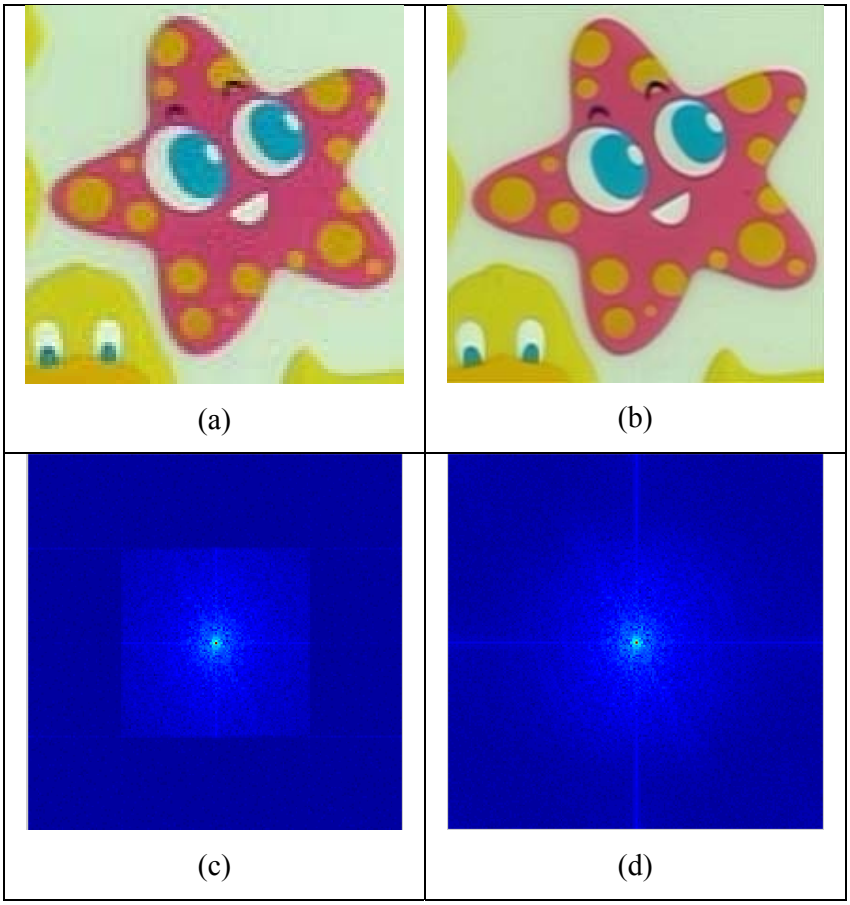


Figure 5-15 – SR through turbulent water currents. Figure a) is the stable reference frame interpolated to twice of its original size, while Figure b) is the super-resolved image. Figures c) and d) show corresponding image spectra intensities of a) and b).

5.5.1.1. Camera Fill Factor

The camera fill factor is the ratio of the active detection area (the size of the light sensitive photodiode) to the inter pixel distance, as shown in Figure 5-16, where a pixel consists of sensitive detection area (white) and insensitive areas (gray), due to the electronics needed for the sensor's operation. The fill factor value is smaller than or equal to 1 (the ratio between the gray area to the entire sensor area). Camera photo detectors introduce low pass filtering to the images captured by the camera. Large fill factor means better light energy efficiency of photo detectors, which in return, results in higher degree of low pass filtering and loss of image high spatial frequencies. Figure 5-17 illustrates frequency responses of photo detectors with different fill

factors. The resolution of images acquired by cameras is ultimately limited by this low pass filtering.

In the evaluation of the results obtained for real-life video, in sections 5.1 and 5.2, one should take into account that substantial resolution enhancement can be expected only if the camera fill-factor is small enough. The camera fill-factor determines the degree of low-pass filtering introduced by the optics of the camera. Due to this low pass filtering; image high frequencies in the base-band and aliasing high frequency components that come into the base-band due to image sampling are suppressed. Those aliasing components can be recovered and returned back to their true frequencies outside the base-band in the described SR process, but only if they have not been lost due to the camera low pass filtering. The larger fill-factor is, the heavier unrecoverable resolution losses will be.

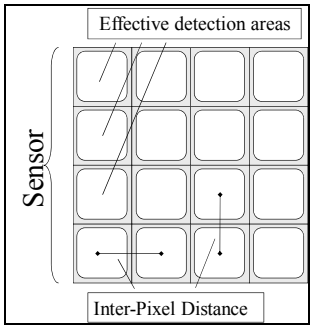


Figure 5-16: Illustration of the fill factor and the inter pixel distance

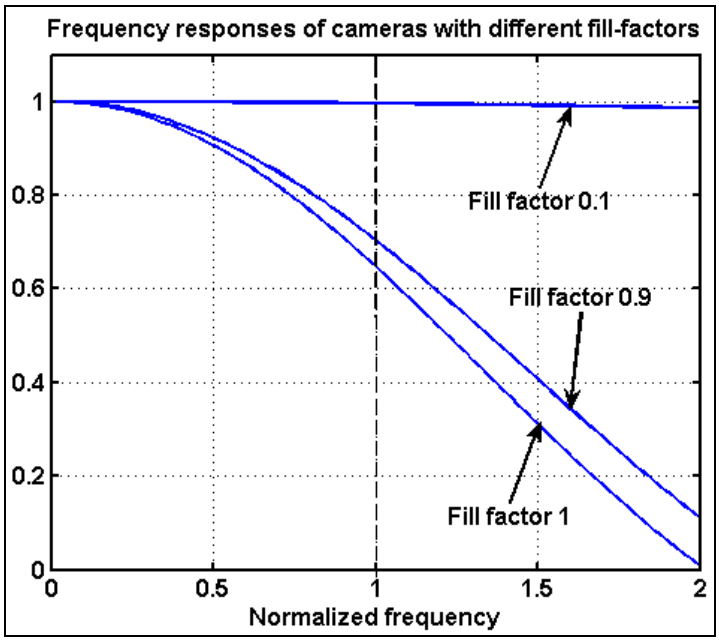


Figure 5-17 – Camera frequency responses for large and small fill factors. Frequency axis is normalized to the width of the camera base band

5.5.1.2. Turbulence Intensity

Pixel displacements due to atmospheric turbulence are chaotic and therefore are characterized by their statistical moments. In our study, the intensity of the turbulence was specified by the standard deviation of the motion vector length. Weak turbulence with low standard deviation of the motion vector magnitude causes small shifts. In this case, the low-resolution (LR) video will virtually appear static and the frames will be nearly identical. Intensive turbulence with standard deviation by the order of the inter-pixel distance may cause substantial aliasing.

5.5.1.3. The number of processed low resolution frames

The quality of the resolution-enhanced frames depends on the amount of data used for their formation. More frames can potentially produce better quality. However, when the number of frames is becoming large enough, having more frames will not necessarily supply more (or significantly more) new pixels, while it will require more processing time.

5.5.1.4. The number of iterations in the process of re-interpolation

Once the motion vectors for each available low resolution frame are known, pixels in the sub-sampled reference frame are replaced with known pixels from those frames. As a result, an up-sampled reference frame that contains pixels (samples) from all the low resolution images is obtained. At this stage it is required finally to perform image re-interpolation to remove aliasing and to generate the best approximation to the image from the given set of pixels. This is achieved by the iterative interpolation algorithm, which converges to the best band limited approximation of the image. Larger number of iterations means that the final image will be closer to the best possible approximation within a given bandwidth. However, iterations consume time, and therefore a compromise should be sought between the resulting image quality and computation time.

5.5.1.5. Creating a Low-Resolution Turbulent Video

The input parameters for the simulation were camera fill factor and the frame-wise pixel translation maps for simulating the turbulence effect. The realizations of the motion vector maps were generated in the form of X- and Y-shift arrays of pseudo-random Gaussian random numbers with a given standard deviation as

described in Sect. 4.5.1. As illustrated by Figure 5-18, for each realization of the pixel translation map, a corresponding low-resolution frame was produced by means of down-sampling of up-sampled low pass filtered high-resolution image according to the sampling grid specified by the corresponding motion vector map.

The up-sampling factor used in the model was M=10, and the down-sampling factor was N=20. Once low resolution frames were obtained, they were used as an input sequences for the SR algorithm. The computer model's out frames are presented in Figure 5-19, where figure (c) is the low resolution turbulence degraded output generated from the original high resolution test image (a), using the motion vector map presented in figure (b). Similarly, Figure 5-20(a) and (b) are, correspondingly, high-resolution text image and its corresponding low-resolution degraded output.

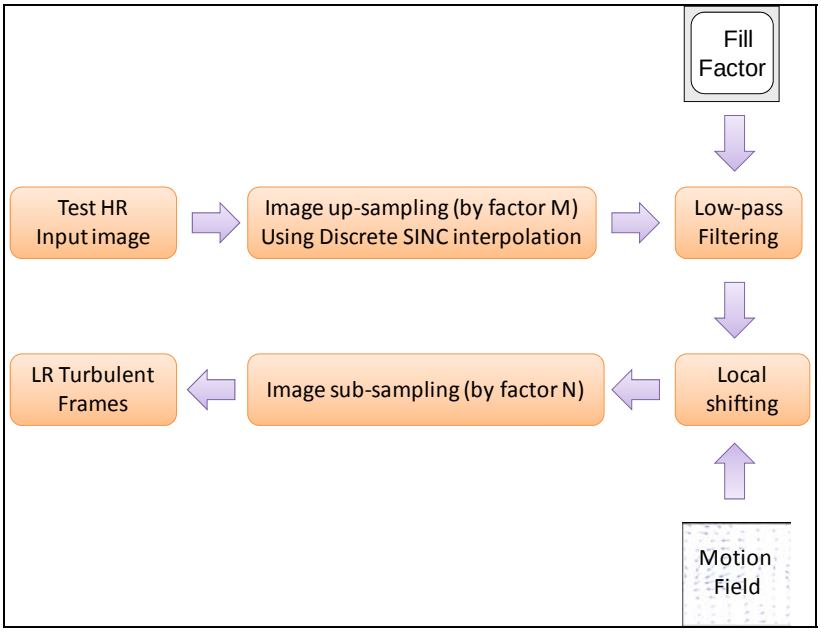


Figure 5-18 Flow diagram of the computer model

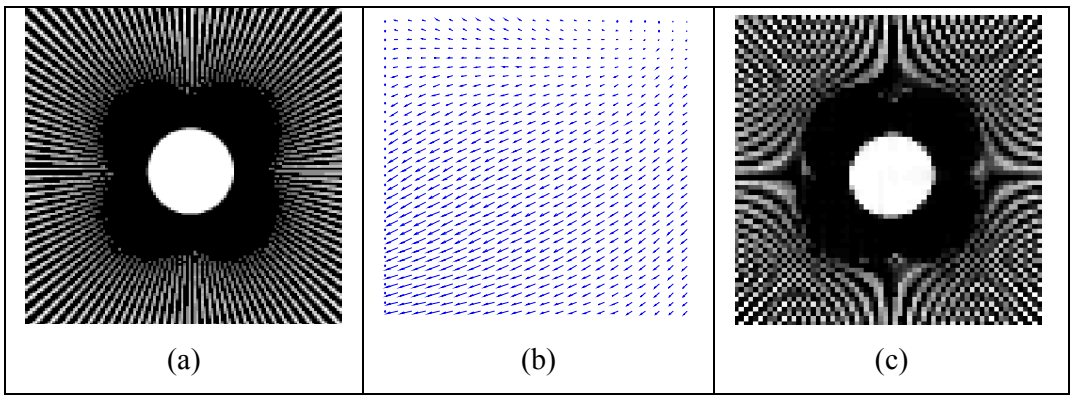


Figure 5-19 – Computer generated test sequence: Figure (c) is the low resolution output generated from figure (a), which is the original high resolution test image, using the motion vector map presented in figure.

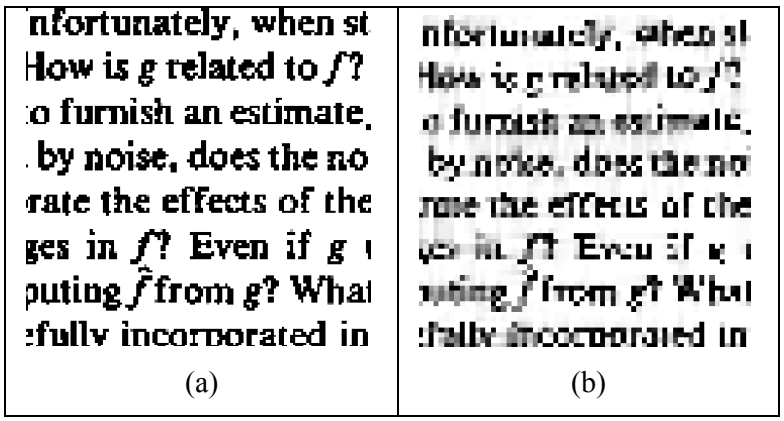


Figure 5-20: (a) Original HR Text Image (b) Turbulent LR Text image

5.5.2. Simulations, Results and Conclusions

This Sect. discusses the SR results obtained when processing the simulated turbulent videos generated by the simulation software. Those results allow evaluating the potential and limitations of the proposed algorithms.

5.5.2.1. Camera Fill Factor

As described in Sect. 5.5.2.1, the camera fill factor determines the degree of low pass filtering of images acquired by the camera. Figure 5-21 illustrates results of generating super-resolved images from sequences of turbulence distorted low-resolution images acquired with cameras with fill-factors 0.05, 0.5 and 0.95. In all cases, the number of low-resolution frames used was 30, the standard deviation of the motion vector length was 0.5 pixels and 50 iterations were used for re-interpolation. It can clearly be seen that as small the fill factor as better the results achieved. Figures (d)-(f) illustrate the same notion through images' spectra and demonstrate that images acquired with larger fill factor have less content in higher frequencies.

5.5.2.2. Turbulence intensity

Results of studying influence of turbulence intensity on the efficiency of image resolution recovery through the SR process are illustrated in Figure 5-22. Super-resolved images shown in Figure 5-22 were obtained from 30 low resolution frames, the camera fill factor was 0.05 and 100 iterations were used in the interpolation. Examining these images, it is evident that turbulence with standard deviation of motion vector lengths of about 0.5 inter-pixel distances creates a sort of optimal conditions for image resolution recovery. Rarely, the physical conditions which affect turbulence can be governed. With that, the camera's zoom can be adjusted so the

motion field's standard deviation becomes 0.5 pixels. Furthermore since the motion field and its characteristics can be computed in real-time, the camera's zoom settings can be adjusted autonomously by the observation system.

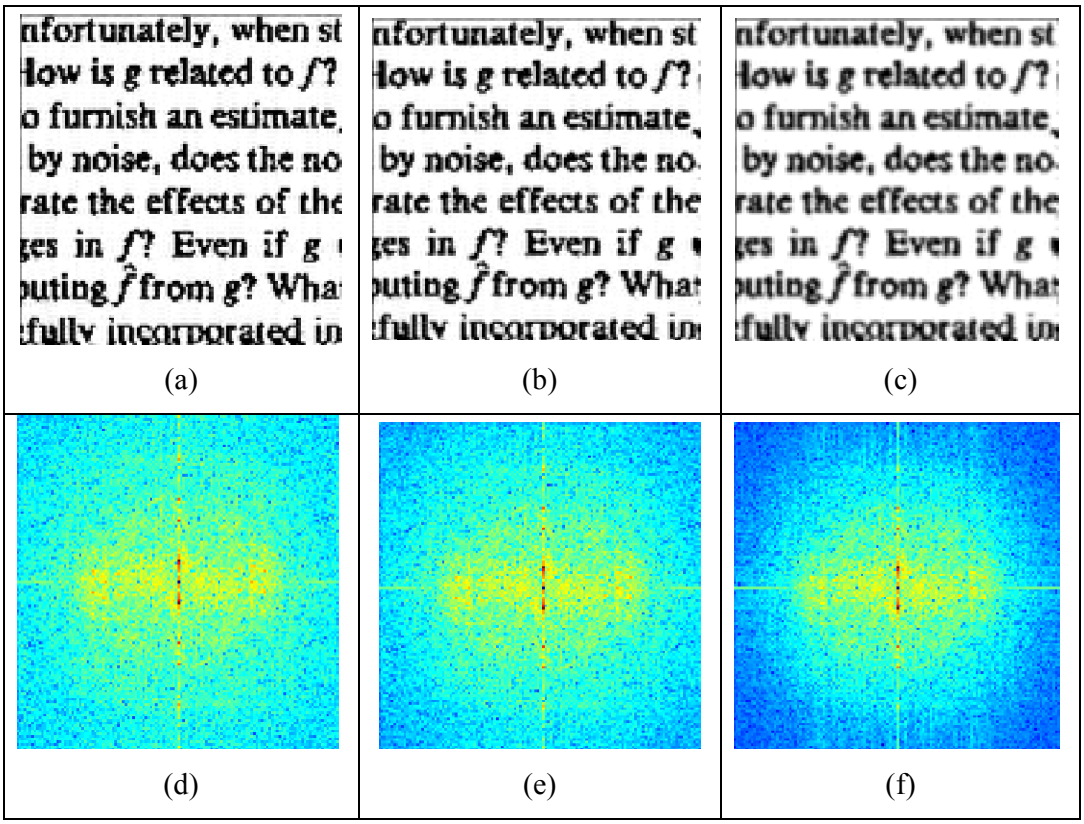


Figure 5-21– Super-resolved images obtained from low resolution images acquired by cameras with different fill factors: a) - fill factor 0.05; b) - fill factor 0.5; c) - fill factor 0.95. Figures (d)-(f) show corresponding image spectra intensities displayed in pseudo colors.

5.5.2.3. The number of processed frames

Obviously, the number of processed low-resolution frames directly affects the SR performance, as more frames provide more additional samples forming denser sample grid. The question, should be addressed with this respect, is how many frames are needed to enable resolution improvement for a given turbulence intensity? Ideally, to obtain two times higher resolution one needs to supply 3 additional samples for each initial low resolution sample which means 3 additional frames for each low resolution frame. Our simulation, however, has shown that in reality the number of additional frames must be much larger. This allows elimination of outliers which might be a result of system noise or abnormal errors in the motion estimation schemes which was integrated as a part of the SR method (see Chapter 3.). This finding is illustrated in Figure 5-23 which presents two test images. In these experiments, camera fill factor

was 0.05, standard deviation of motion vector length was 0.5 and 50 interpolation iterations were used.

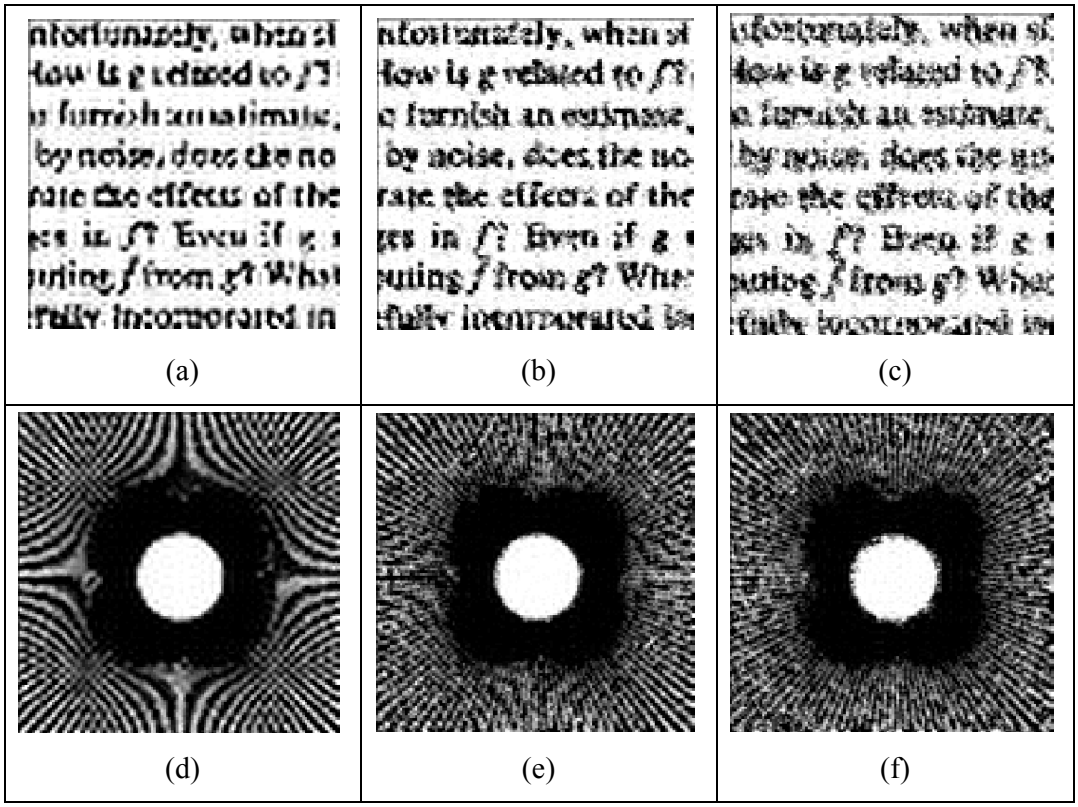


Figure 5-22 – SR results obtained from low resolution images distorted by atmospheric turbulence with different intensity: (a) standard deviation (STD) of 0.1 inter-pixel distance; (b) STD of 0.4 inter-pixel distance; and (c) STD of 0.8 of inter-pixel distance. (d) Motion vectors magnitudes' standard deviation of 0.1 inter-pixel distance; (e) STD 0.45 of inter-pixel distance; and (f) STD 0.9 of inter-pixel distance.

5.5.2.4. The number of iterations in the process of re-interpolation

Image re-interpolation is the final stage of the SR process aimed at recovery of those samples in the dense sampling grid that were not obtained from the accumulated low resolution frames. It is implemented through an iterative interpolation algorithm, which converges to the best band limited approximation of the image. The latter claim is justified and dealt with, thoroughly, in the succeeding chapter.

Figure 5-24 (a) through (c) show how the number of iterations influences the quality of final super-resolved image. In this experiment, camera fill factor was 0.05, standard deviation of vector motion lengths was 0.5 of inter-pixel distances and 30 low resolution frames were used. Figure 5-25 illustrates the iteration process. It shows

a typical dependence of the energy of the difference between subsequent images in course of iterations from the number of interpolation iterations. From these figures one can see that the number of iterations is a quite critical parameter of the restoration process and that, to achieve a good restoration quality, one needs about 100 iterations. Figure 5-25 shows the convergence of the process.

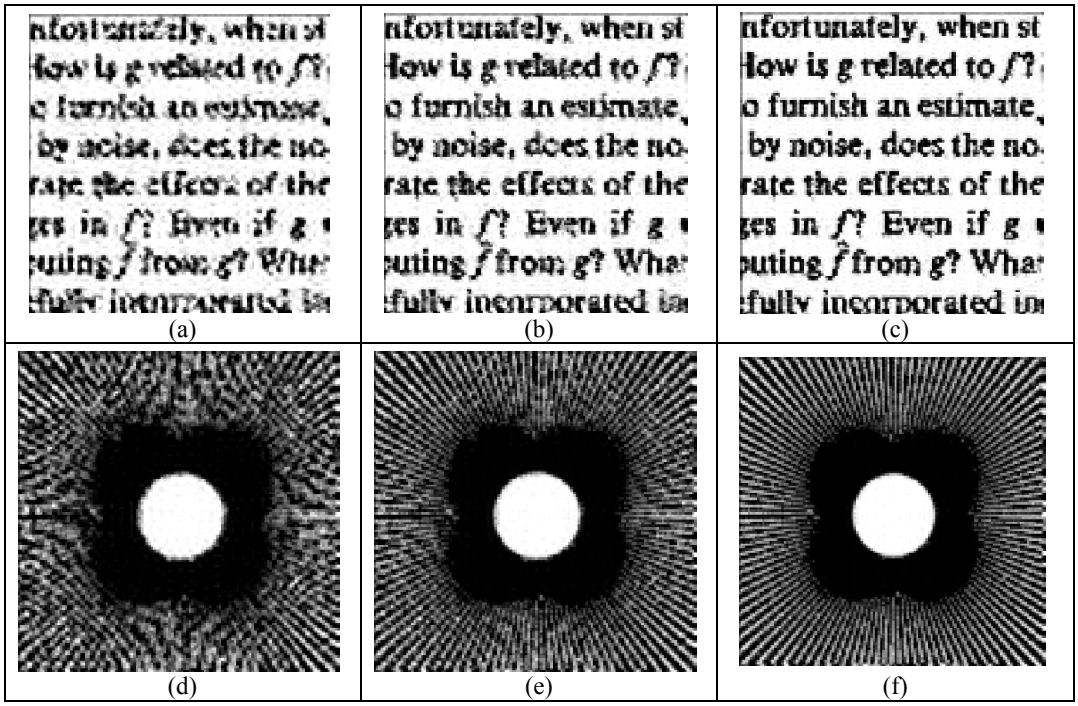


Figure 5-23 - Results of image resolution enhancement from 5 (figures (a) and (d)), 15 ((b) and (d)), and 30 ((c) and (f)) low resolution turbulent images.

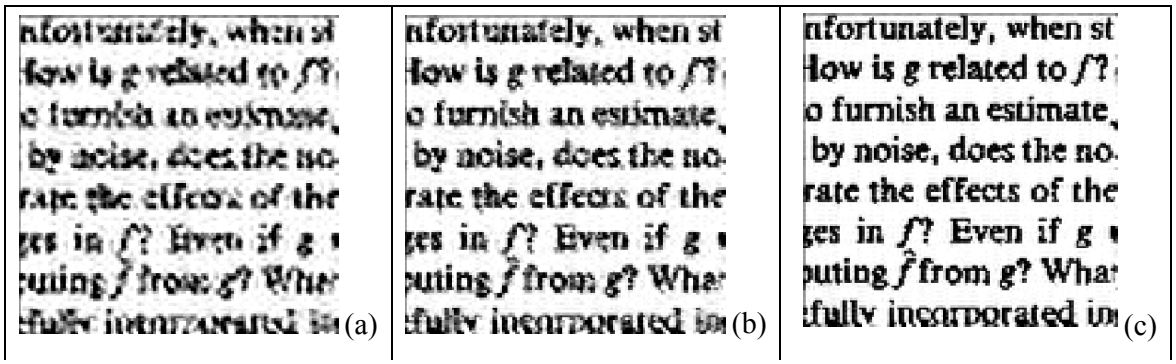


Figure 5-24– The SR process as a function of the number of iterations of the re-interpolation scheme. The number of iterations used to compute figures (a), (b) and (c) are 5, 20 and 100 iterations, respectively.

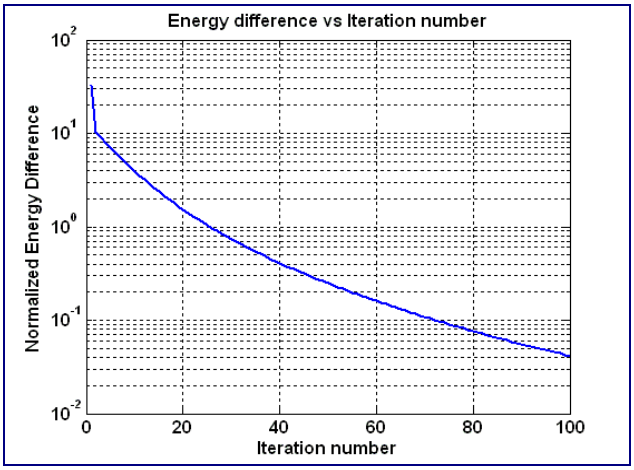


Figure 5-25: Energy of the difference between subsequent images in course of iterations as a function of the number of interpolation iterations

5.6. Super-resolution in Real-time

Nowadays most digital footage data is transmitted and stored using the International Telecommunication Union (ITU) and Moving Picture Experts Group (MPEG) coding standards [130,131]. SR techniques that have been designed for raw, uncompressed, video may not be effective when applied to compressed video because they do not incorporate the compression process into their models. This has raised the need for SR techniques which utilize the standards' various features. A real-time SR method, over video encoder hardware, was presented in [147]. Yet, the suggested method utilized proprietary compression standard and exploited only the sequence's motion field. Using a common compression standard, such as ITU H.264 or MPEG-4 presents several benefits, such as real-time compatibility and broad availability of software and hardware implementations. In this section a practical SR scheme, which utilizes MPEG-4 features for producing, in real-time, good quality higher-resolution videos is described [148].

5.6.1. Super-resolution within the MPEG-4 Framework

MPEG-4 Part 10 or MPEG-4 AVC (for Advanced Video Coding) is a standard for video compression [149]. It is also known as, H.264. The intent of H.264/AVC was to create a standard capable of providing good video quality at substantially lower bit rates, without increasing the complexity of design so much that it would be impractical or excessively expensive to implement. An additional goal was to provide enough flexibility to allow the standard to be applied to a wide variety of applications

on a wide variety of networks and systems, including low and high bit rates, low and high resolution video, broadcast, DVD storage, RTP/IP packet networks, and ITU-T multimedia telephony systems.

A digital image sequence coded at a low bit rate using a motion-compensated video compression standard should contain little data redundancy. However, the success of a particular SR enhancement algorithm is predicated on sub-pixel-resolution overlap (i.e., redundancy) of moving objects from frame-to frame. If an MPEG bit stream is coded at a relatively high bit rate (e.g., a compression ratio of 15:1), enough data redundancy exists within the bit stream to successfully perform SR enhancement, as a part of the decoding process held by the video decoding hardware.

5.6.1.1. Video Object Coding

One of the key contributions of MPEG-4Visual is a move away from the 'traditional' view of a video sequence as being merely a collection of rectangular frames of video. Instead, MPEG-4 Visual treats a video sequence as a collection of one or more video objects (VO). MPEG-4 Visual defines a video object as a flexible 'entity that a user is allowed to access (seek, browse) and manipulate (cut and paste) [150]. An instance of a VO at a particular point in time is a *video object plane* (VOP). A VOP is defined by its texture (luminance and chrominance values) and its shape. In general, MPEG-4 coding of a VOP involves coding of motion, texture, and shape information. To enable access to an arbitrarily shaped object, such an object has to be separated from the background and the other objects.

MPEG-4 video object coding consists of shape coding (for arbitrarily shaped VOs), motion compensated prediction to reduce temporal redundancies, and DCT-based texture coding of the motion compensated prediction error data to reduce spatial redundancies. The video coding is performed at the macroblock level. VOPs are divided into macroblocks, such that they are represented with the minimum number of macroblocks within a bounding rectangle. Similar to MPEG- 1 and MPEG-2, MPEG-4 supports intracoded (I), temporally predicted (P), and bidirectionally predicted (B) VOPs, all of which are described in the following section.

5.6.1.2. MPEG Picture Types

MPEG-encoded image sequences are divided into groups of pictures (GOPs) composed of primarily three different frame types: intra-coded frames (I-pictures), predictive-coded frames (P-pictures), and bi-directionally predictive-coded frames (B-

pictures). I-pictures are coded independently without reference to other pictures, using block-DCT (discrete cosine transform) compression. P-pictures obtain predictions from temporally preceding I- or P-pictures, while B-pictures are predicted from the nearest preceding and/or upcoming I- or P-pictures.

In the video observation model, the I-picture which begins a particular GOP is the reference frame. The SR algorithm requires motion vectors between the reference frame and all other frames. Therefore, only the neighboring frames up to and including the first P-picture in the current GOP and the frames down to and excluding the last P-picture in the previous GOP are integrated with the I-picture.

5.6.1.3. Computing the Reference Frame

The stable estimation of the scene is computed through pixel-wise temporal median (see 4.2). Since all motion vectors are computed with respect to the last I-frame, the suggested solution takes the first I-frame since the following frames have predictions which are directly connected to macro-blocks within the I-picture. Integration of more than one GOP into the SR process is desired, therefore, for every new I-frame the translations with regards to the reference I-frame have to be computed.

5.6.1.4. Motion analysis

Estimating accurate sub-pixel-resolution motion vectors is a critically important component in SR enhancement algorithms. The goal is to compute the motion fields as quickly and as accurately as possible, so that SR enhancement becomes practical desktop computing application. As described in Sect. 5.1, the SR process requires fractional translations rather than just integer values. The MPEG-4 standard defines half-pixel vectors in MPEG-4 Simple Profile and quarter-pixel vectors in Advanced Simple profile and H.264. Sub-pixel motion estimation requires the encoder to interpolate between integer sample positions in the reference frame. Interpolation is computationally intensive. Calculating sub-pixel samples for the entire search window is not usually necessary. Instead, it is sufficient to find the best integer-pixel match (using one of the fast search algorithms discussed in 3.2) and then to search interpolated positions adjacent to this position. In the case of quarter-pixel motion estimation, first the best integer match is found; then the best half-pixel position match in the immediate neighborhood is calculated; finally the best quarter-pixel match around this half-pixel position is found.

5.6.1.5. Real Motion Extraction

The pixel displacement map is analyzed and segmented to separate pixels of the real moving objects from those that belong to the stable scene and are displaced solely due to the atmosphere turbulence. Using the MPEG-4 scene segmentation for different VOP's, a video scene may be made up of a background object and a number of separate foreground objects. The separate objects may be exploited for efficient background separation. One drawback, however, is the fact that the MPEG-4 standard was designed for natural videos, therefore the presence of turbulent motion might lead to a significant number of errors in the process of segmenting the scene into VO's. This results in more areas that are regarded as real moving objects (see 4.3) and are omitted from the data accumulation process. This is coped by accumulating more low resolution frames in the process.

5.6.1.6. Global Motion Extraction

Macro blocks within the same video object may experience similar motion. For example, camera pan will produce apparent linear movement of the entire scene, camera zoom or rotation will produce a more complex apparent motion and macro blocks within a large object may all move in the same direction. Global Motion Compensation (GMC) enables an encoder to transmit a small number of motion (warping) parameters that describe a default 'global' motion. Additionally, the GMC can provide improved motion analysis and real motion extraction.

5.6.2. Generation of super-resolved stable frames

Once the data accumulation and interpolation stages are done, the output-stabilized and resolution-enhanced image obtained might be subjected to additional processing aimed at camera aperture correction, denoising and reducing blocking (de-blocking) and ringing effects (de-ringing). The goal of a de-blocking or de-ringing filter is to minimize the effect of blocking or ringing distortion whilst preserving important features of the image. MPEG-4 Visual describes a deblocking filter and a deringing filter: these are 'informative' parts of the standard and are therefore optional. Both filters are designed to be placed at the output of the decoder. With this type of post-filter, unfiltered decoded frames are used as the reference for motion-compensated reconstruction of further frames. This means that the filters improve visual quality at the decoder but have no effect on the encoding and decoding processes.

5.6.3. Results

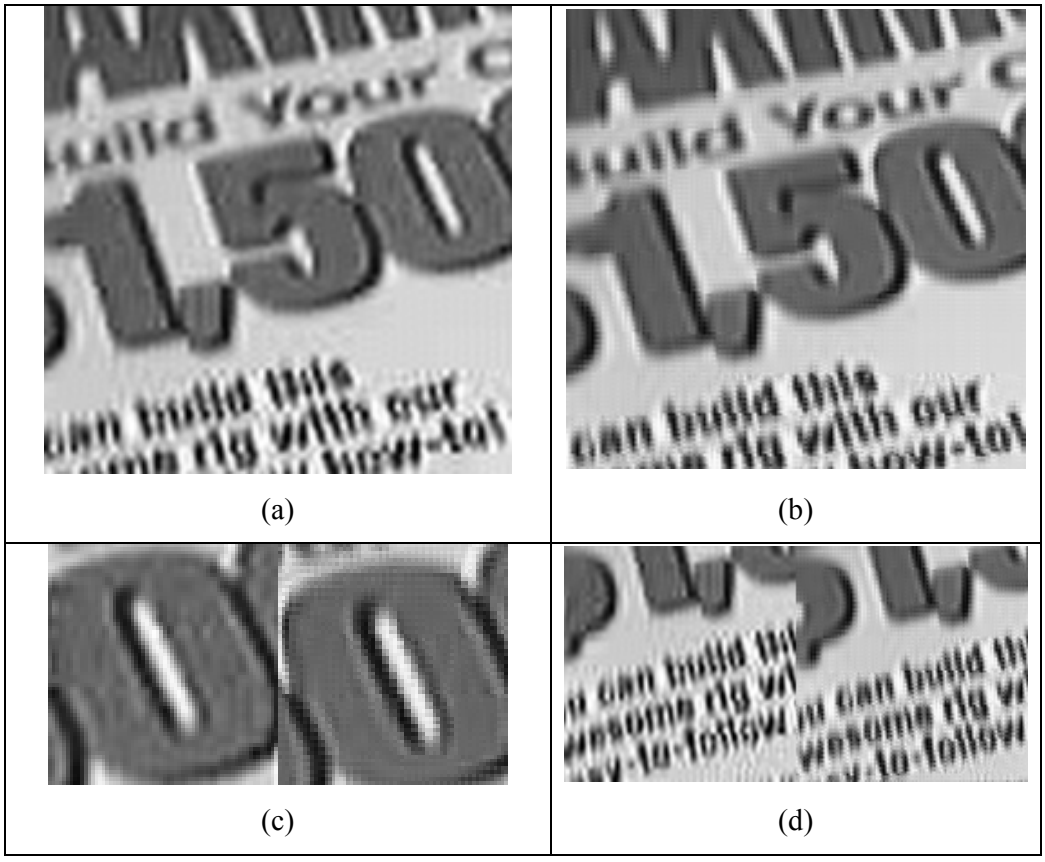
Figure 5-26 and Figure 5-27 present experimental results, in which decoded pictures from MPEG-4 bit streams containing translations due to global and turbulent motions. Figure 5-26 depicts the SR process, through MPEG-4, applied on a real-life sequence containing global motion. The sequence was acquired with 320x240 pixels webcam. Figure 5-26(a) presents the first frame extracted from the sequence interpolated to twice, on each axis, of its original size. Figure 5-26(b) is the super-resolved image after accumulation of 50 frames. The translation of each frame with regards to the first frame, which is the first I-frame broadcasted, was computed using the MPEG-4 diamond search with quarter pixel accuracy. Figure 5-26(c) and (d) present fragments extracted from figures (a) and (b) respectively. The left-hand side parts of (c) and (d) are interpolated (Figure 5-26 (a)), while the right-hand parts are the corresponding super-resolved fragments.

A super-resolved frame computed using 150 frames of a turbulent degraded sequence is presented in Figure 5-27(a). Figure (b) depicts the absolute difference of the super-resolved image and the reference frame interpolated to four times of its original size. As one can see, most of the difference's energy is located in the vicinity of edges. The right-hand side of Figure 5-27(c) shows a fragment extracted from the interpolated reference frame, while the left-hand side is the corresponding super-resolved one. As one can see the later contains finer details of the scene. Finally, Figure 5-27(d) presents both spectra of the interpolated (on the left-hand side of Figure 5-27(d)) and of the super-resolved frames. It can be seen that the SR process enhance data in higher frequencies.

5.7. Summary

The chapter has shown theoretically and practically that distortion caused by turbulence or camera motions can be used to increase image resolution beyond camera's limitation. In order to obtain good results, it is required to have some motion in the scene to allow sub-pixel re-sampling. Loosely speaking, the more frames in the given video the more information is given and therefore better results can be obtained. However, at some stage more frames will not supply new information and therefore the SR results will reach saturation. Several dozens of low-correlated frames are required to have significant results. Another limitation over the SR performance is the fill factor of the camera. Large fill factor causes to a loss of high frequencies. Those

frequencies are lost and cannot be restored. A good interpolation method is also important, to remove aliasing of the sampling, and to achieve the best approximation to the original high resolution image from the given samples. The interpolation methods and the recovery of missing data on a non-uniform sampling is the subject of Chapter 6. For evaluating the super-resolved images a quantitative *image effective bandwidth (IEBW)* measure is introduced. This measure corresponds to the human eye quality perception and is invariant to the image's size, hence the quality of images with different sizes can be compared.



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

Figure 5-26 – SR from Global Motions through MPEG-4

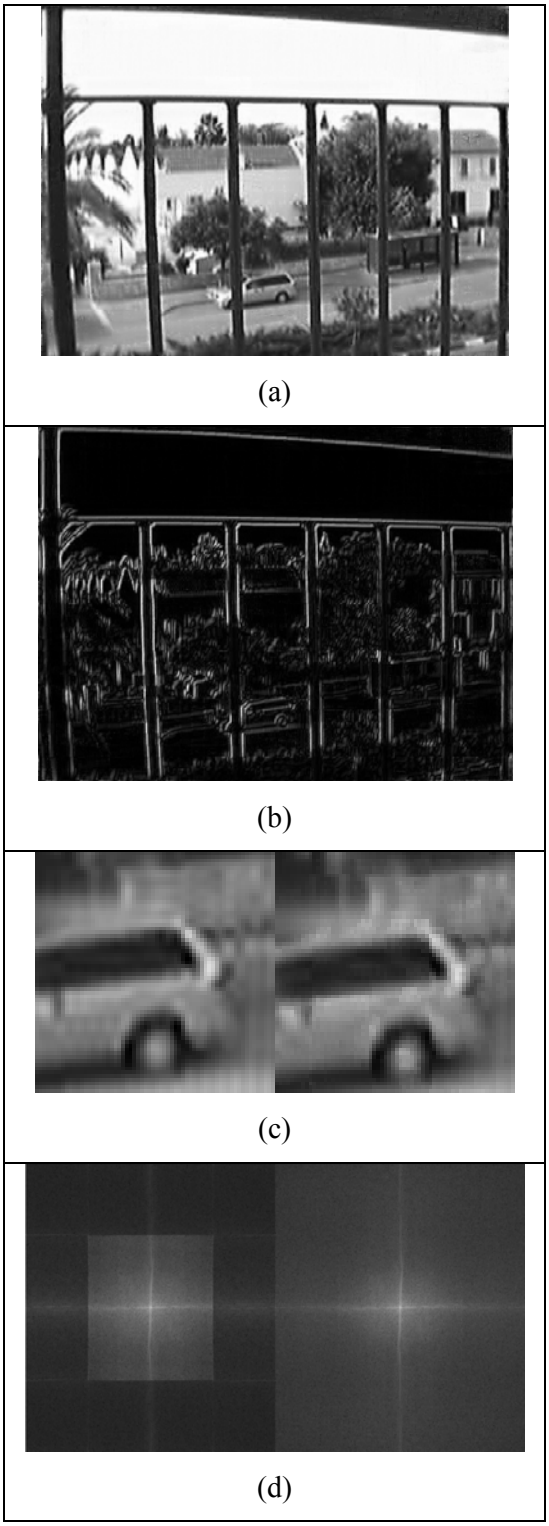


Figure 5-27 - SR of turbulent degraded videos MPEG-4.

6. Sparse Data Interpolation

In many applications it may happen that sampled data are collected in irregular fashion and/or it may occur that some samples of a regular sampling grid are lost or unavailable. Various sparse data interpolation algorithms are known. This chapter unifies known algorithms and addresses the problem of signal reconstruction from irregular samples and recovery of missing data as equivalent tasks and suggests a new unified approach to optimal recovery based on the Discrete Sampling Theorem, which is formulated within its scope. This theorem provides a tool for optimal, in terms of root mean squared error, approximation of arbitrary discrete signals specified by their sparse samples, provided appropriate selection of the signal representation transform.

6.1. Preface

The discrete sampling theorem states that band-limited discrete signals can be precisely recovered from their sparse samples, provided positions of the available samples satisfy certain limitations. While each of the previous works in the field (see Sect. 1.4), makes a major algorithmic contribution, which contains within it key ingredients that could be used to establish the discrete sampling theorem, we believe that suggested framework is a coherent and offers the most natural way to define the sampling theorem for digital data, and which gives certain algorithmic ideas an intellectually clear role in supporting that framework. The applicability of the framework is presented for both known sparse data reconstruction methods, based on the discrete Fourier and Cosine transforms [151], and for new ones, based on the Radon, Haar, Walsh and Wavelets transforms.

The rest of the chapter is organized as follows, in Sect. 6.3, the mathematical foundations, of this approach, are provided by the discrete sampling theorem. In Sect. 6.4 we discuss the validity of the assumptions put in the base of the presented approach. In Sect. 6.5, the properties of certain transforms, which are specifically relevant for signal recovery from sparse data, are discussed. Finally, in Sect. 6.6 we provide an experimental evidence of precise signal reconstruction from sparse data. Additionally, possible applications are addressed in Sect. 6.7.

6.2. Preliminaries

For the formulation of the discrete sampling theorem, we assume that:

- There are $\boldsymbol{K}$ samples of a continuous signal, taken at irregular positions within a certain known interval, $\boldsymbol{X}$, of the signal support.
- The physical coordinates of the available samples are known with a certain accuracy, $\boldsymbol{\Delta x}$. The ratio of the position accuracy and the support interval define the signal regular sampling grid with the number of sampling positions, where the $\boldsymbol{K}$ available signal samples are sparsely scattered over this sampling grid.

$$N = \frac{X}{\Delta x} > K \tag{6-1}$$

- The $\boldsymbol{N}$ samples constructing the discrete signals, which represent continuous ones, are "band-limited" in a domain of a certain discrete orthogonal transform.
- The goal of the processing is generating out of this incomplete set of $\boldsymbol{K}$ samples a complete set of $\boldsymbol{N}$ samples specified in each node of the grid.

6.3. The Discrete Sampling Theorem Formulation

Let A_N be a vector of N samples $\{a_k\}_{k=0,\ldots,N-1}$ which completely define a signal. Let Q be a K-size proper subset of {0,1,..,N-1} and assume that only the $K < N$ samples $\{a_k\}_{k\in Q}$ are known. Let Φ_N be an NxN orthonormal (unitary for the complex case) matrix. The columns of Φ_N ($\{\varphi_r\}_{r=0,\ldots N-1}$) define an orthonormal basis of R^N. It follows that

$$\mathrm{A}_N = \Phi_N^T \Gamma_N, \quad \left\{ a_k = \sum_{m=0}^{N-1} \gamma_m \varphi_k(m) \right\}_{k=0,\ldots,N-1}, \tag{6-2}$$

where Γ_N is the signal's transform coefficients ($\{\gamma_r\}_{r=0,\ldots,N-1}$) vector when representing A_N by the basis vectors $\{\varphi_r\}_{r=0,\ldots N-1}$. Consequently, the available $\boldsymbol{K}$ signal samples define a system of $\boldsymbol{K}$ equations:

$$\left\{ a_k = \sum_{m=0}^{N-1} \gamma_m \varphi_k(m) \right\}_{k\in Q}. \tag{6-3}$$

Let $\boldsymbol{R}$ be a P-size proper subset of {0,1,..,N-1} where P<K. A discrete band-limited signal approximation to the signal A_N is obtained by

$$\left\{ \hat{a}_k = \sum_{m\in R} \gamma_m \varphi_k(m) \right\}_{k=0,\ldots,N-1} \tag{6-4}$$

The underlying approximation model is the assumption that all transform coefficients with indices $r \notin R$ are equal to zero. Assuming that A_N is a discrete band-limited signal (i.e., $a_k = \hat{a}_k$, $_{k=0,\ldots,N-1}$), the available signal samples $\{a_k\}_{k\in Q}$ can now be expressed as:

$$\left\{ a_k = \sum_{m\in R} \gamma_m \varphi_k(m) \right\}_{k\in Q} \tag{6-5}$$

Eq. 2.5 is, in fact, a linear system consisting of $\boldsymbol{K}$ equations with $P \le K$ unknowns. Existence and uniqueness of a solution to this linear system of equations is dependent, for any transform, on the subsets $\boldsymbol{Q}$ and $\boldsymbol{R}$.

For any signal $\mathbf{A}_N$, by virtue of the Parseval's theorem, the band-limited signal $\hat{\mathbf{A}}_N$ approximates $\mathbf{A}_N$ with a mean squared error

$$MSE = \left\| A_N - \hat{A}_N \right\| = \sum_{k=0}^{N-1} \left| a_k - \hat{a}_k \right|^2 = \sum_{r\notin R} \left| \gamma_r \right|^2 \tag{6-6}$$

This error can be minimized by an appropriate selection of ortohonormal basis defined by the transform Φ_N and the subset $\boldsymbol{R}$ of used vector basis. In order to do so, one must know the transform $\boldsymbol{\Phi}_N$ basis functions energy compaction ordering. If one knows, for a class of signals, a transform that features the best energy compaction in the smallest number of transform coefficients, one can, by selection of this transform, secure the best band-limited approximation of the signal $\{a_k\}$ from the given subset $\{\tilde{a}_k\}$ of its samples.

In this way, we arrive at the following Discrete Sampling Theorem, which is formulated in the following two statements:

Statement 1. *For any discrete signal of* $\boldsymbol{N}$ *samples defined by its* $\boldsymbol{K \le N}$ *sparse and not necessarily regularly arranged samples, its band-limited, in terms of certain transform* $\boldsymbol{\Phi}_N$*, approximation can be obtained with mean square error defined by Eq.* (6-6)*. The approximation error can be minimized by using a transform with the best energy compaction property.*

Statement 2. *Any signal of* N *samples, that is known to have only* $K \le N$ *non-zero transform coefficients for a certain transform*

Φ_N (discrete band-limited signal), can be fully recovered from its subset of K samples, as long as Eq. 2.5 is solvable.

6.4. Validity of the assumptions

The applicability of the above derivation depends on the validity of the assumption that band-limited, in certain basis, approximation of signals is an acceptable solution in data recovery. To this end, there is a trivial necessary condition, that $P \le K$, hence the number of samples should be at least equal to the dimension of the sub-space of the band-limited signals non-zero transform coefficients. A validation to this assumption has been presented in signal compression algorithms. In data compression applications, such transforms as DCT and certain wavelets are known for their very good energy compaction properties for wide variety of signals in image and audio processing and are successfully used for compression by means of replacement of signals by their low pass band-limited approximations [152]. Recent advances in compressive sensing [153] also support signal band-limitedness assumption. An important application, where this assumption is supported by a physical reality is computed tomography where slice projections can very frequently be regarded as band-limited, in Radon transform, signals as the outer parts of slices are usually known to be empty.

6.5. Various Transforms Analysis

6.5.1. Discrete Fourier and Cosine Transforms

Let Φ_N be the NxN Discrete Fourier transform matrix and $\widetilde{A}_k$ are DFT band-limited signals of $\boldsymbol{N}$ samples with only $\boldsymbol{K}$ nonzero low frequency DFT coefficients. The discrete signal A_N can be precisely recovered from exactly $\boldsymbol{K}$ of its arbitrarily taken samples, if $\boldsymbol{R}$ is a contiguous set (modulo N) in the transform domain.

As it follows from the theorem's statements and from Eqs. (6-4)-(6-6), the theorem is proven if matrix $\mathbf{KofN}_{\boldsymbol{DFT}}^{\boldsymbol{LP}}$ is invertible. While a succinct proof is presented here, the complete proof and discussion, for the DFT case, can be found in [151].

Consider the $\mathbf{K}of\mathbf{N}_{DFT}^{LP}$-trimmed $\boldsymbol{DFT_N}$ matrix, given by:

$$\mathbf{K}of\mathbf{N}_{DFT}^{LP} = \left\{ \exp\left(i2\pi \frac{kr}{N} \right) \right\}_{k \in Q} \tag{6-7}$$

Without loss of generality, we can assume that the indices of the given contiguous samples are $\{a_{k_1}, a_{k_2}, \cdots, a_{k_P}\}$ and the signal's bandwidth is between the frequencies $(-M, -M+1..., M-1, M)$. Substituting in equation (6-7) gives:

$$\begin{bmatrix} a_{k_1} \\ a_{k_2} \\ \cdots \\ a_{k_P} \end{bmatrix} = \begin{bmatrix} \exp\left(\frac{i2\pi k_1(-M)}{N}\right) & \exp\left(\frac{i2\pi k_1(-M+1)}{N}\right) & \cdots & \exp\left(\frac{i2\pi k_1(M)}{N}\right) \\ \exp\left(\frac{i2\pi k_2(-M)}{N}\right) & \exp\left(\frac{i2\pi k_2(-M+1)}{N}\right) & \cdots & \exp\left(\frac{i2\pi k_2(M)}{N}\right) \\ \cdots & \cdots & \cdots & \cdots \\ \exp\left(\frac{i2\pi k_p(-M)}{N}\right) & \exp\left(\frac{i2\pi k_p(-M+1)}{N}\right) & \cdots & \exp\left(\frac{i2\pi k_p(M)}{N}\right) \end{bmatrix} \begin{bmatrix} \alpha_{-M} \\ \alpha_{-M+1} \\ \cdots \\ \alpha_{M} \end{bmatrix} \tag{6-8}$$

Showing that $\mathbf{K}of\mathbf{N}_{DFT}^{LP}$ is a nonsingular matrix, can be found by calculating its determinant in the following way:

$$\det KofN = \exp\left(\frac{i2\pi M \sum_{t=1}^{p} k_t}{N} \right) \det \begin{bmatrix} 1 & \exp\left(\frac{i2\pi k_1}{N}\right) & \cdots & \exp\left(\frac{i2\pi k_1 2M}{N}\right) \\ 1 & \exp\left(\frac{i2\pi k_2}{N}\right) & \cdots & \exp\left(\frac{i2\pi k_2 2M}{N}\right) \\ \cdots & \cdots & \cdots & \cdots \\ 1 & \exp\left(\frac{i2\pi k_p}{N}\right) & \cdots & \exp\left(\frac{i2\pi k_p 2M}{N}\right) \end{bmatrix} \tag{6-9}$$

The first matrix in the product of matrices, described in Eq. (6-9), is a diagonal, which is obviously invertible. The second one is a version of Vandermonde matrices, which are also known to have non-zero determinant if, like in our case, its ratios for each row are distinct [154], so that $\det KofN \neq 0$ and therefore $rank(KofN) = p$ for every set of ***K*** samples.

The theorem was proven for any arbitrary ***K*** sample positions and therefore suggests that any band-limited signal, in DFT domain, can be recovered from it sparse representation, given that a sufficient contiguous number of samples is given. Sufficient, as explained earlier, is the number of coefficients in the DFT domain.

Now, Let Φ_N be the NxN Discrete Cosine transform matrix and $\widetilde{A}_k$ are DCT band-limited signals, in DCT domain, of ***N*** samples with only ***K*** nonzero low frequency DCT coefficients. The discrete signal A_N can be precisely recovered from exactly ***K*** of its arbitrarily taken samples.

N-point Discrete Cosine Transform of a signal is equivalent to 2***N***-point Shifted Discrete Fourier Transform (SDFT) with shift parameters (1/2,0) of 2***N***- sample signal obtained from the initial one by its mirror reflection [155]. The shifted DFT matrix is simply a product of a DFT matrix and a nonsingular diagonal matrix. Hence, the proof for DCT is essentially the same as in the DFT case.

These theorems also hold for the two-dimensional (2D) DFT and DCT transforms, provided that the band-limitation conditions are separable [156]. The case of non-separable band-limitation requires further study. However, in Sect. 6 we provide experimental results that show that precise recovery of 2D signals with circularly band-limited DCT spectrum is possible for virtually arbitrary positions of its sparse samples. Note that working in DFT or DCT domain, in the case of low-pass band-limitation, results in signal discrete sinc-interpolation [118].

6.5.2. The Discrete Radon Transform

A straightforward application the discussed sparse data recovery algorithm can find in tomographic imaging, where it frequently happens that a substantial part of slices, which surrounds the body slice, is known to be an empty field. This means that slice projections (sinograms) are Radon transform "band-limited" functions. Therefore whatever number of projections is available, a certain number of additional projections, commensurable, according to the discrete sampling theorem, with the size of the slice empty zone, can be obtained and the corresponding resolution increase in the reconstructed images can be achieved within the described framework. Another option is recovery of projection data that might be missing due to sensor faults or to other reasons.

In order to describe the sparse data recovery through the Radon transform within the suggested discrete sampling theorem, one must have a discrete transform and its algebraically exact inverse. While the continuum theory defines the Radon transform and its inverse, the discrete equivalent is not a trivial problem.

Some authors have attempted to exploit the projection-slice theorem [157]. In the continuum theory, this says that Radon transform can be obtained by (a) performing a 2-d Fourier Transform, (b) obtaining a radial slice of the Fourier transform, and (c) applying a 1-d inverse Fourier transform to the obtained slice. This suggests an algorithm for discrete data, by replacing steps (a) and (c) by discrete fast Fourier transforms for data on 2-d and 1-d Cartesian grids, respectively. However, strictly

speaking, this continuum approach is problematic since step (b) is not naturally defined on digital data: the 2-d FFT outputs data in a Cartesian format, while the radial slices of the Fourier domain typically do not intersect the Cartesian grid. Therefore, some sort of interpolation is required, and so the transform is not algebraically exact. Also, even if the transform should turn out to be invertible (which may be very difficult to determine) the transform is typically not invertible by any straightforward algorithm.

Other authors have attempted to exploit two-scale relations, which say that if one knows the Radon transform over four dyadic sub-squares of a dyadic square these can be combined to obtain the Radon transform over the larger square [158,159,160,161]. This suggests a recursive algorithm, in which the problem is broken up to the problem of computing Radon transforms over squares of smaller sizes which are then recombined. Strictly speaking, however, the driving identity is a fact about the continuum and does not directly apply to digital arrays, so that when this principle is operationalized, the results involve interpolation and other approximations, and end up being quite crude compared to what we have in mind here. Finally, the use of two-scale relations means that summation along lines is approximated by summation along line segments which are not exactly parallel and so the results can lack a certain degree of geometric fidelity.

Recently, a stable forward and inverse Radon transform called Fast Slant-Stack [162] was suggested. The suggested transform scheme, which coaligns with the continuum theory, presents algebraic exactness, geometric fidelity, fast computation and invertibility. Those characteristics make this Slant-Stack discrete Radon transform most suitable for sparse data recovery within the suggested framework.

The Slant-Stack transforms images, with size $(N \times N)$, to a corresponding Radon image with size of $(2N \times 2N)$. Let $\boldsymbol{I}$ denote the vector space of $(N \times N)$ arrays and $\boldsymbol{D}$ denote the vector space of $(2N \times 2N)$ arrays. The transform $\Re : I \to D$ and its inverse can be written in the following matrices formulation:

$$Y = \Re X \tag{6-10}$$

$$X = \Re^{\#} Y \tag{6-11}$$

where $Y \in D$, $X \in I$ and $\Re^{\#}$ is the generalized inverse applied to an array Y.

The transform matrix $\Re$ size is $\left(N^2 \times 4N^2\right)$ and its rank is N^2 [162]. Consequentially it can be sown that, provided appropriate subset of samples of $\boldsymbol{D}$, the transformation matrix full rank is preserved and the truncated transformation matrix, KofN_{Radon} has an inverse. With that, this does not mean that there is an algorithm that can actually invert it. For that to happen, one need the mapping to have a small condition number [163].

6.5.3. Wavelets and other bases

The main property of wavelet bases is that their basis functions are most naturally ordered in terms of two components: scale and position within the scale. Scale index is analogous to the frequency index for DFT. Position index tells only of the shift of the same basis function within the signal on each scale. Therefore band-limitation is analogous to scale limitation for wavelets. Limitations, in terms of position, are trivial: it simply means that some parts of the signal are not relevant. Commonly, discrete wavelets are designed for signals whose length is an integer power of 2 ($N = 2^n$). For such signals, there are $s \le n$ scales and possible band-limitations.

The simplest form of wavelet bases is the Haar basis functions. The one dimensional Haar wavelet's mother function, $\psi(\tau)$, can be described as:

$$\psi(t) = \begin{cases} 1, & 0 \le t < \frac{1}{2} \\ -1, & \frac{1}{2} \le t < 1 \\ 0, & otherwise \end{cases} \tag{6-12}$$

and its scaling function $\varphi(t)$ can be described as:

$$\phi(t) = \begin{cases} 1, & 0 \le t < 1 \\ 0, & otherwise \end{cases} \tag{6-13}$$

A signal with $\boldsymbol{N = 2^n}$ samples and band-limitation of P, on the Haar domain (the transform coefficients $\{P+1,...,N\}$ are zero) is composed of, at most, P piecewise constant intervals. For any two samples that are located on the same interval, the representation in the Haar domain is the same. Therefore, having more than one sample per constant interval will not change the rank of the matrix KofN. The condition for perfect reconstruction is to have at least one sample on each of those intervals (all in all at least P samples for the KofN matrix).

For the Walsh basis functions, the index corresponds to the "sequency", or to the number of zero crossings of the basis function. The sequency carries a certain analogy to the signal frequency. Basis functions ordering according to their sequency, which is characteristic for Walsh transform, preserves, for many real signals, the property of decaying transform coefficients' energy with their index. Therefore, for Walsh transform the notion of low-pass band-limited signal approximation, similar to the one described in Sect. 6.5.1, for DFT, can be used. On the other hand, Walsh basis functions can be characterized by the scale index similarly to that of Haar basis functions. A signal with $N = 2^n$ samples and band-limitation of P has $2^{\lceil \log_2 P \rceil}$ equal length intervals, in which each basis function has a constant value. For example, if $N = 1024$ and $P = 5$, then there are 8 intervals, each of length of 128 samples. In order to have perfect reconstruction, it is needed to have P samples taken from different intervals. Unlike the Haar case, not all the intervals are needed to be sampled, but only P intervals out off the total number of intervals.

For the special case in which P is a power of 2, then there are P intervals in which each one of them needs to be sampled, which means that in this special case, the reconstruction condition of Walsh is identical to Haar. However, while have similar condition for reconstruction, the reconstruction process itself is different in those two cases.

6.6. Experimental Verification and Possible Applications

6.6.1. Introductory

Signal recovery, from sampled data, through direct matrix inversion, is, generally, a very demanding computational procedure. In applications, where one can be satisfied with signal reconstruction with limited accuracy, simple iterative reconstruction methods can be applied [151]. In the following examples, where iterative algorithm is used, the Gerchberg-Papoulis's procedure [135,136] was utilized.

For the experimental verification of the validity of the discrete sampling theorem we used real-life test images and pseudo-random band-limited 1D and 2D test signals with uniform spectrum and a selected bandwidth. For the generation of those signals, the scheme, which was described in Sect. 3.4 for DCT band-limited signals, is

extended to any arbitrary transform. Hence Eq. (3-10) can be written in the following form:

$$I_{test}(\alpha) = T^{-1}\{T\{I_{seed}\} \cdot LPF\ Mask(\alpha)\} \tag{6-14}$$

6.6.2. Exact signal recovery from sparse samples

In addition to the experimental evidences that can be found in the literature for DFT, DCT and Wavelets [55,60,151] cases, in this section, we provide evidence of the possibility of exact recovery of band-limited signals from sparse data. Figure 6-1 - illustrates exact reconstruction, by the direct matrix inversion and by the iterative algorithm, of a band-limited signal of 64 samples from its 13 samples taken at random positions or concentrated within of a certain interval of signal support. From the graph of signal approximation error versus the number of iterations of the iterative algorithm, one can see that the speed of convergence of the iterative algorithm might be quite low. Experiments show that the convergence speed of the algorithm depends on the spread of the available samples: the more uniformly they are spread over the signal support interval the higher the convergence speed.

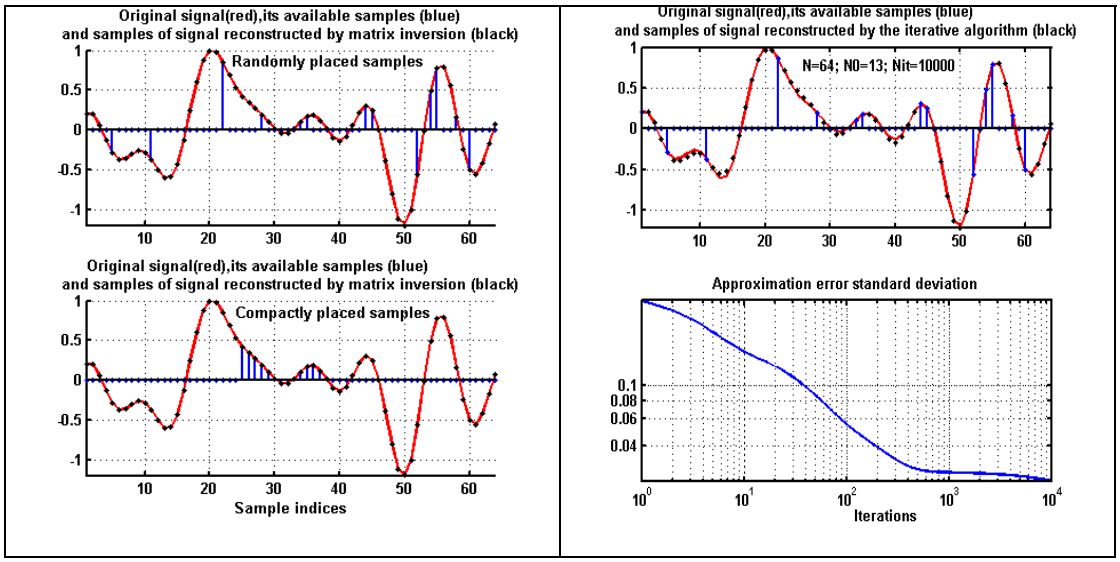


Figure 6-1 - Restoration of a DFT low pass band-limited signal by matrix inversion for the cases of random (upper left) and compactly placed signal samples (bottom left) and by the iterative algorithm (upper right). Bottom right plot shows standard deviation of the signal restoration error as a function of the number of iterations. The experiment was conducted for test signal of 64 samples; with bandwidth of 13 frequency samples (~1/5 of the signal base band).

The image presented in Figure 6-2, is a 256x256 pixel test image low-pass band-limited in DCT domain by a 90^{o} circle sector to have only 3217 non-zero spectral

coefficients out of 256x256. The image was sampled in 3217 positions shown in Figure 6-2, a) with white dots and then reconstructed from these sparse samples using the iterative reconstruction algorithm with the reconstruction error $\mathbf{3.86 \cdot 10^{-8}}$ achieved after 500 iterations. The result of the reconstruction is shown in Figure 6-2, b). Note that in this case the iterative algorithm converges reasonably fast as it can be concluded from the graph of standard deviation of the reconstruction error as a function of the number of iterations shown in Figure 6-2, c).

While the iterative algorithm is capable of precise reconstruction of band-limited signals, Spline interpolation, for this set of signals, can only produce an approximated output. Standard deviation of the reconstruction error achieved using the Wolberg at el. Spline interpolation algorithm [57] for this test image is shown for comparison in the graph of Figure 6-2, c) by a dotted line. The reconstruction error by itself is shown in Figure 6-2, d). For the implementation of the multilevel B-Splines algorithm a code kindly provided by Wolberg was used.

Figure 6-3 - illustrates a 2D signals band-limited in the Haar transform domain. Two examples are illustrated: arrangement of sparse samples for which signal recovery is possible and that for which signal is not recoverable. Following the theory, outlined in Sect. 6.5.2, when the Haar reconstruction is possible, it is trivial, hence the nearest neighbor interpolation method is used.

A perfect reconstruction of Walsh band-limited signal is depicted in Figure 6-4, where the $\boldsymbol{N}$=512 and $\boldsymbol{P}$=5. In this example, the resulted $KofN^{Walsh}$ obtained matrix is:

$$KofN_{p=5}^{Walsh} = 0.0442 \begin{bmatrix} 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 \end{bmatrix} \tag{6-15}$$

and its rank equals to 5. One should note, in this specific example, perfect reconstruction is not possible in the Haar transform domain since one of the intervals contains no samples.

6.7. Applications

There might be numerous applications of above described algorithms for recovery band-limited signals and generating band-limited approximations of signals

from sparse data. In this section we briefly review few of them. For the latter, the above theory and algorithms can be applied as following:

1. Given a certain number of available signal samples, specify, according to the accuracy with which physical coordinates are known, the signal dense sampling grid and the required number of samples to be recovered.
2. Select a transform with presumably better energy compaction capability for the signal at hand and specify the signal band limitation in the domain of this transform.
3. Place available signal samples on the signal dense sampling grid and run the direct matrix inversion or iterative reconstruction algorithm.

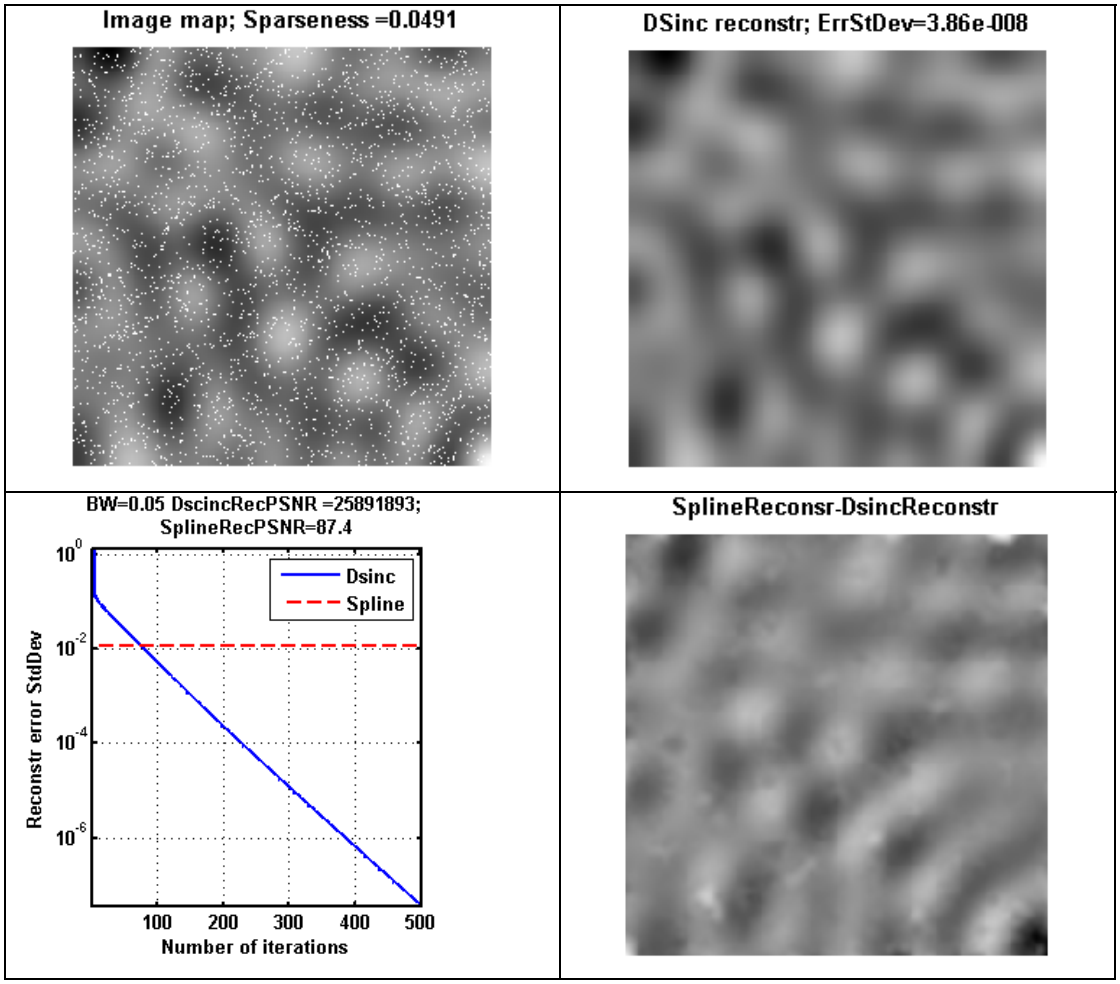


Figure 6-2 - Iterative reconstruction of a band-limited profile versus spline-reconstruction: Figure a) is the initial profile with samples positions shown in white points; b) presents the reconstruction result of the iterative algorithm after 500 iterations; c) depicts the graph of the reconstruction error standard deviation versus the number of iterations (solid line) and the spline reconstruction error (dash line);

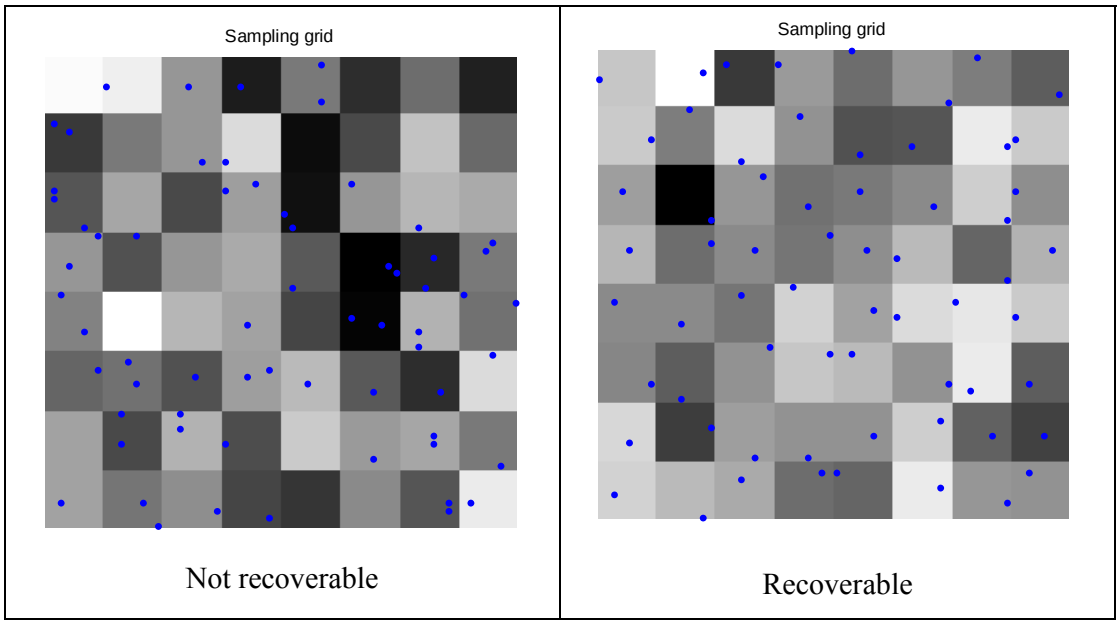


Figure 6-3 - Two cases of sparse sampling of an image band-limited in Haar Transform: a) not recoverable case; b) recoverable case (sample points are marked with dots). Image size was 64x64, and band-limitation was 8x8 (scale 3)

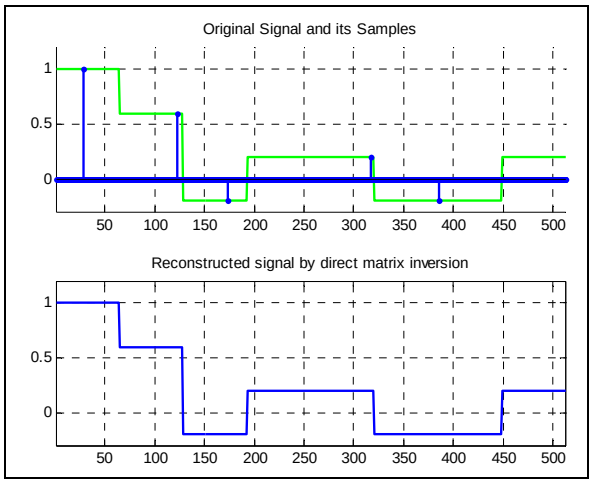
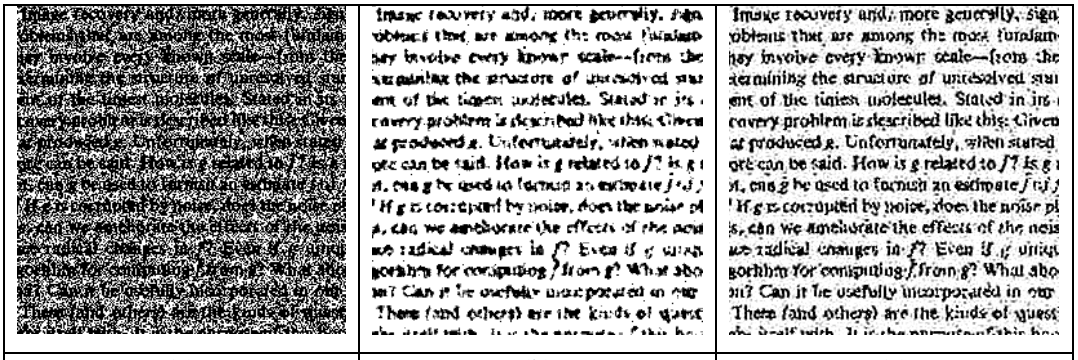


Figure 6-4 - Example for perfect reconstruction on Walsh domain

We illustrate possible applications on several examples. The first application example is recovery of an image corrupted by "salt & pepper" noise. In case of the "salt & papper" noise, image pixels are replaced, with a certain probability of error, with their extreme values that correspond to signal minimum or maximum. With certain probability of false alarms, erroneous pixels can be quite easily detected, and the distorted image can be subjected to the above described iterative reconstruction

procedure that will generate a band-limited approximation of the image that preserves available not distorted pixels, the band limitation being determined by the rate of non-distorted pixels. Figure 6-5 illustrates an example of such restoration for the case when the probability of a missing pixel is 0.5. The original image is presented in Figure 6-5, a). The restoration results using modified median filter over 3x3 window is presented in Figure 6-5, b). The median filter is modified so it incorporates into the restoration only samples with valid data, as those are the only samples incorporated into the interpolation process. The resulting image is depicted in Figure 6-5,c). The image is the interpolation approximation of the original image that is low pass band-limited in DCT domain by a 90^o circle sector of radius 0.7 (in the units of the base band).



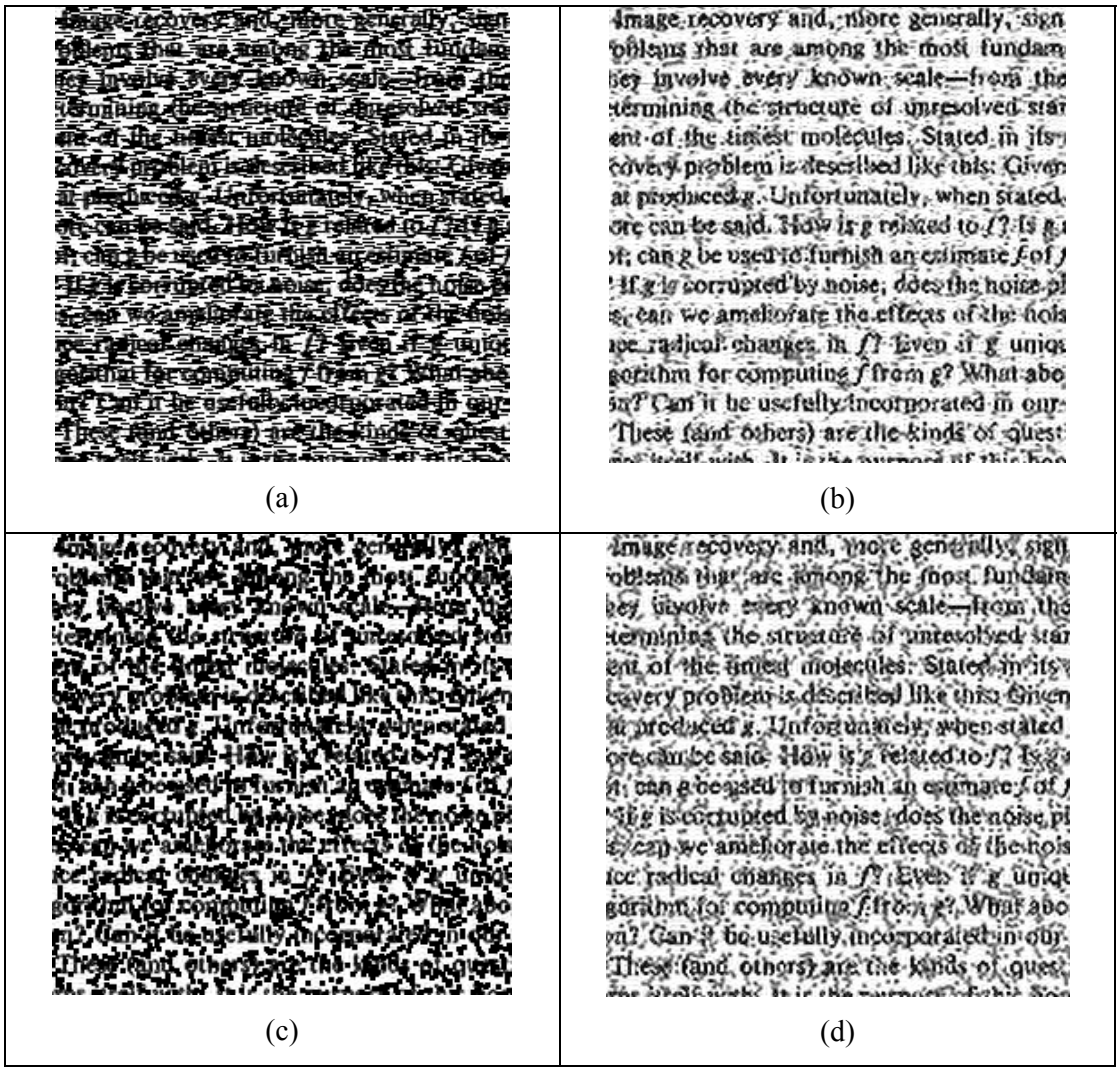
(a) (b) (c)

Figure 6-5 - Recovery of images corrupted by "salt & pepper" noise with probability of pixel missing equal to 0.5.

Certain recovery of missing image data is possible in even more complicated situations when random missing data happens in groups of pixels as it is illustrated in Figure 6-6, a) and b) for the case of 1x3 pixel stripes and in Figure 6-6, c) and d) for the case of 3x3 squares. In both cases the rate of missing pixels was 0.5 and the reconstruction algorithm generated a low-pass DCT domain band-limited approximation to the original image similar to that of Figure 6-5, b).

The main problem in such applications is reliable detection of missing pixels. "Salt & pepper" noise, a general impulse noise and group data losses may occur in digital communication systems used for image transmission. In these cases, error-detecting encoding may be used for detecting missing pixels, which will provide information on coordinates of distorted pixels. There are also applications where positions of missing samples are known from the data acquisition procedure. A typical example is profile function restoration from its level lines. This application is

illustrated in Figure 6-7 that shows that perfect reconstruction of a band-limited 2D profile function is possible from quite sparse level lines that contain only 5% of profile samples.



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

Figure 6-6 - Recovery of images corrupted by chaotic signal omissions in groups of 1x3 pixels (a and b) and groups of 3x3 pixels (c and d). Left column: corrupted images; right column: images restored by the iterative algorithm working in DCT domain.

Yet another potential application of the above signal recovery technique is image super-resolution from multiple video frames with chaotic pixel displacements due to atmospheric turbulence, camera instability or similar random factors [134]. By means of elastic registration of sequence of frames of the same scene, one can determine, for each image frame and with sub-pixel accuracy, pixel displacements caused by random acquisition factors. Using these data, a synthetic fused image can be generated by placing pixels from all available video frames in their proper positions on the

correspondingly denser sampling grid according to their found displacements. In this process, some pixel positions on the denser sampling grid will remain unoccupied, especially, when limited number of image frames is fused. These missing pixels can then be restored using the above described iterative band-limited interpolation algorithm. Computer simulation reported in [93] showed that application of the iterative interpolation may substantially improve results of image resolution enhancement by fusing multiple frames with different local displacements. This is illustrated in Figure 6-8, which shows one low resolution frame (a), image fused from 30 frames (b) and a result of iterative interpolation (c) achieved after 50 iterations. Image band limitation was set in this experiment twice of the base band of raw low resolution images.

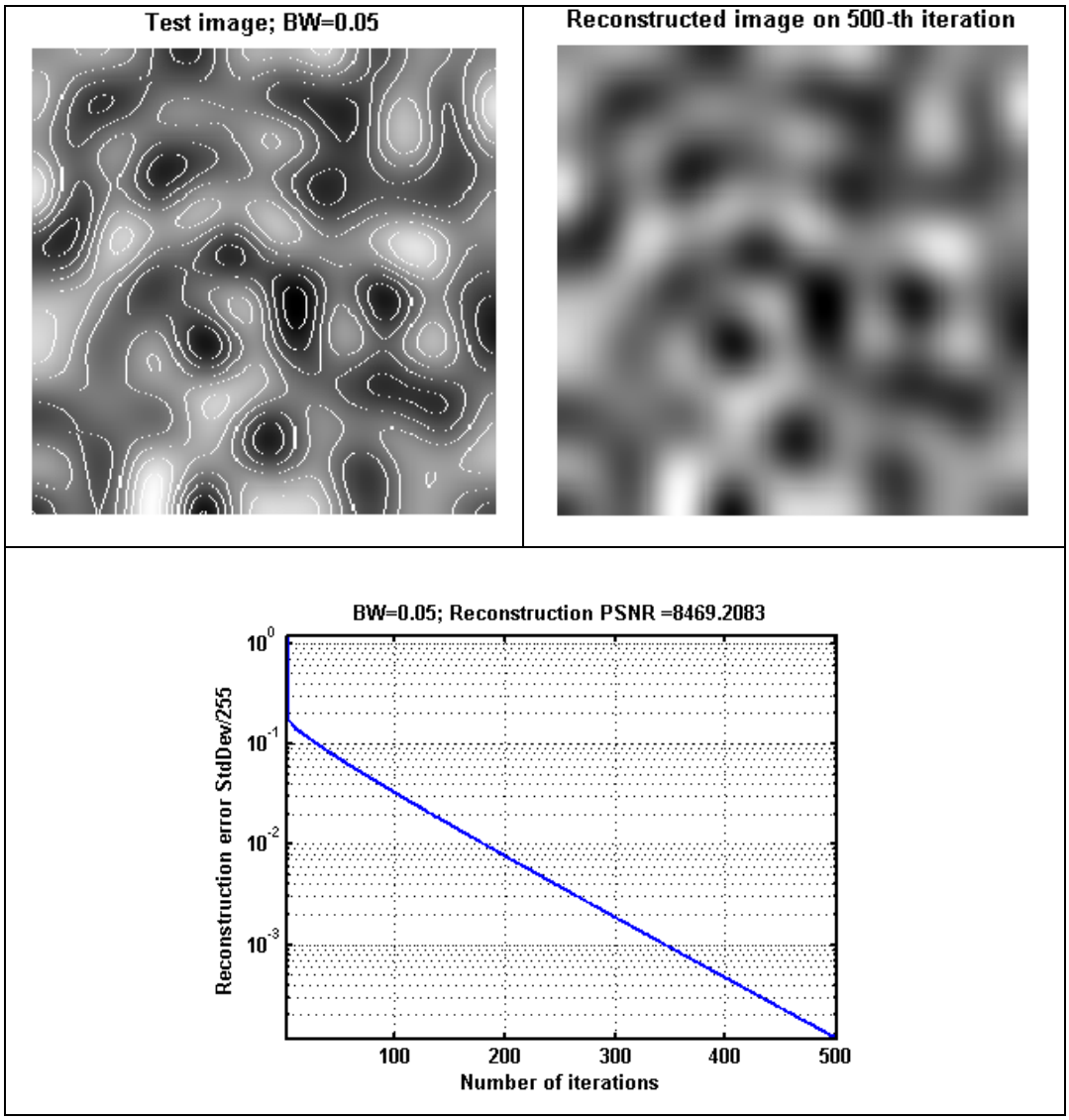


Figure 6-7 - 2D band-limited profile function perfect reconstruction from level lines

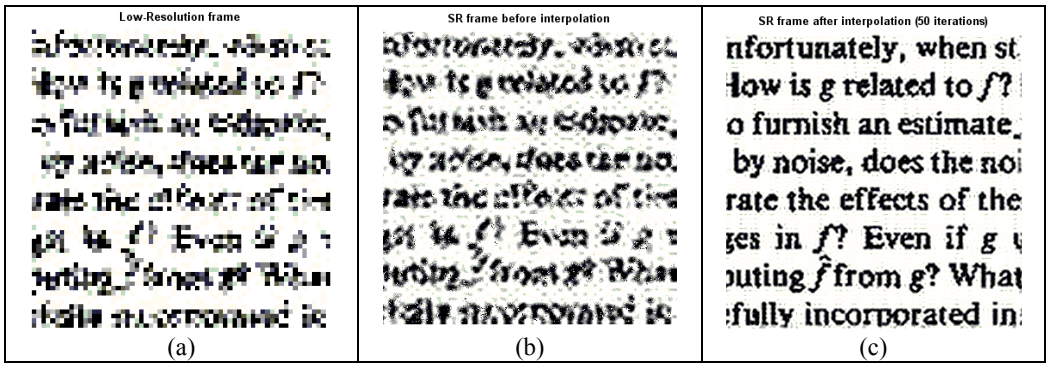


Figure 6-8 - Iterative image interpolation in the super-resolution process: a) – a low resolution frame; b) image fused by elastic image registration from 50 frames; c) – a result of iterative interpolation of image b) after 50 iterations.

As described in Sect. 6.5.2 the exact forward and inverse Radon transform allow the utilization of the discrete Radon transform within the suggested framework. This is illustrated in Figure 6-9 and Figure 6-10, which show that virtually perfect recovery of missing 55% samples of sinograms is possible. Figure 6-9 illustrates recovery of missing projections. In this case the standard deviation of the reconstruction error is not as low as in the previous case, which, perhaps, can be attributed to not full reversibility of the truncated Radon Transforms. Note that fully invertible direct and inverse discrete Radon Transforms used in this experiment were implemented using the code found in [164].

6.8. Summary

This chapter has addressed the problem of signal reconstruction from irregular samples and recovery of missing data. Considering that positions of available signal samples are always specified with a certain accuracy that defines maximal number of signal samples sufficient for signal representation, we treat this problems as equivalent tasks and suggest a new approach to optimal recovery of signals from irregularly sampled or sparse data based on the Discrete Sampling Theorem introduced in Sect. 2. The discrete sampling theorem refers to discrete signals band-limited in a domain of a certain transform and states that "***KofN*** band-limited" discrete signals of ***N*** samples can be precisely recovered from their ***K*** sparse samples provided positions of the available samples satisfy certain limitations, which depend on the transform. This theorem provides also a tool for optimal, in terms of root mean

squared error, approximation of arbitrary discrete signals specified by their sparse samples, with "***KofN***- band-limited" signals, provided appropriate selection of the signal representation transform. Properties of different transforms, such as Discrete Fourier, Discrete Cosine, Radon, Haar, Walsh and wavelet transforms, relevant to application of the Discrete Sampling Theorem are discussed. Finally, different applications are presented of the algorithms for recovery images corrupted by impulse noise, or specified by level lines, for image super-resolution from multiple frames and for tomographic image reconstruction from sparse or sparsely sampled projections.

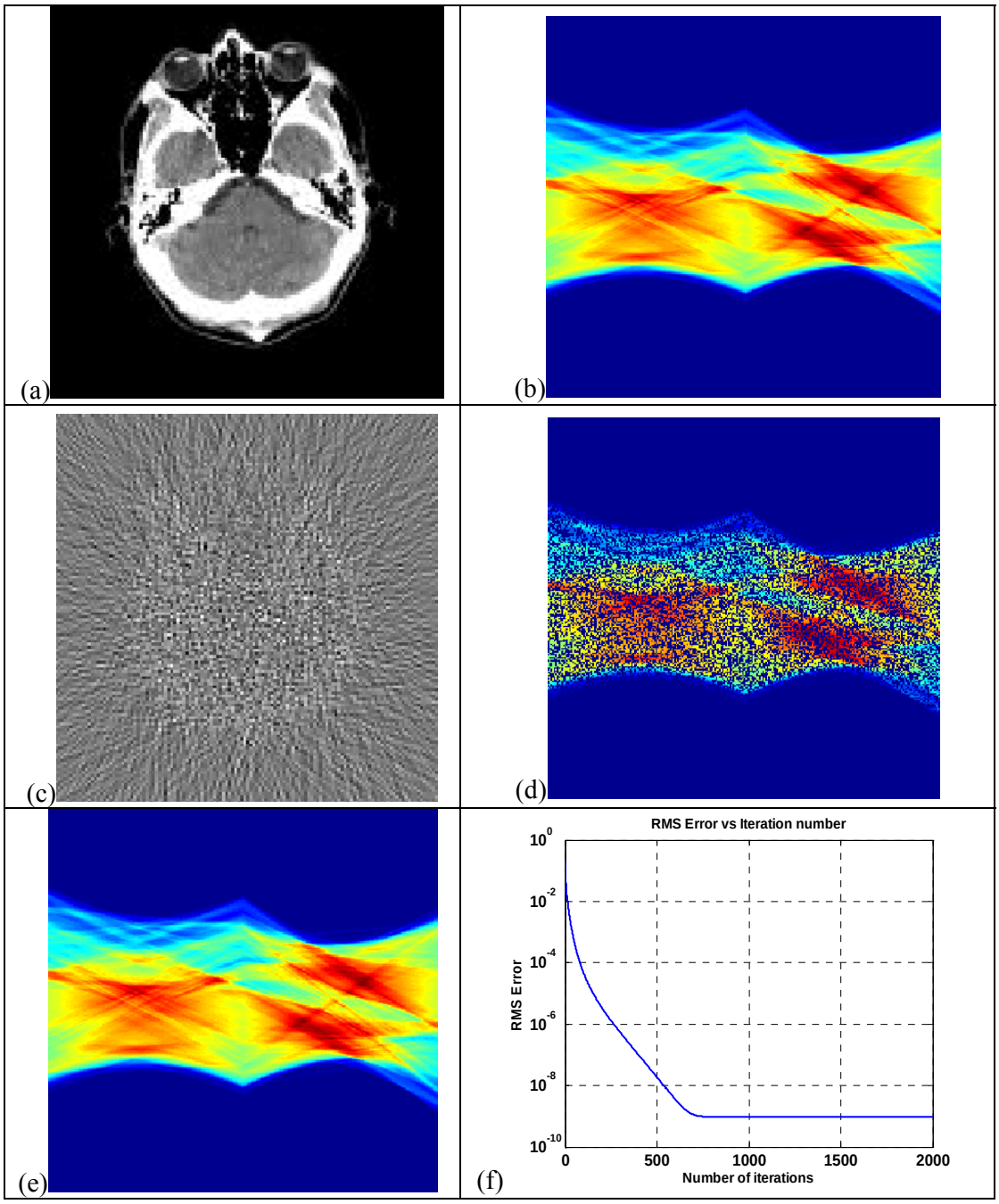


Figure 6-9 - Recovery of missing samples of a sinogram: (a), (b) original image and its Radon transform (sinogram), (c) image reconstructed from the sinogram (d) corrupted by the loss of 55% of its randomly selected samples; e) a sinogram recovered from (d) using the iterative band-limited interpolation algorithm and (f) a plot of standard deviation of slice reconstruction error as a function of the iteration number.

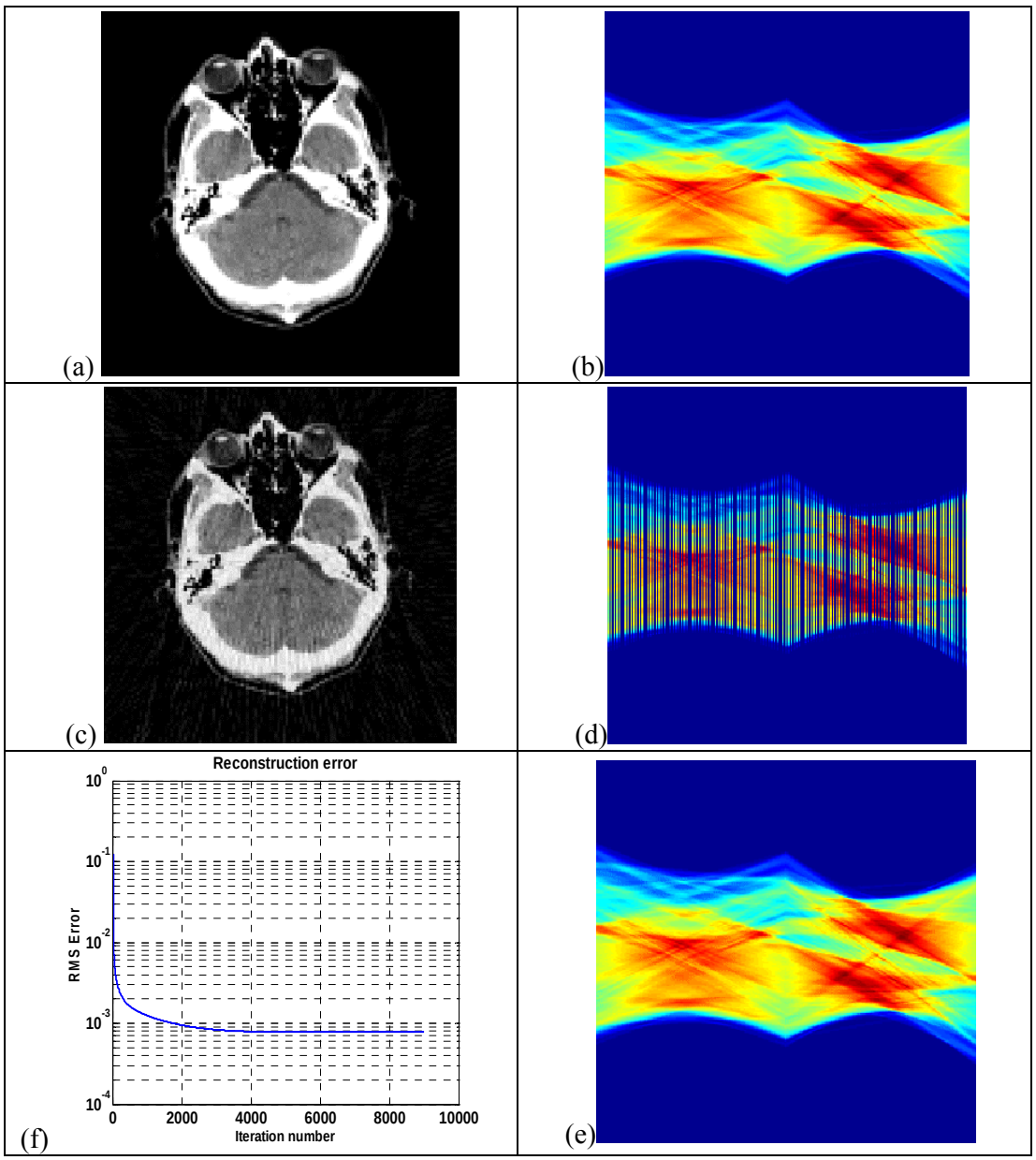


Figure 6-10 - Recovery of missing projections: (a), (b) original image and its Radon transform (sinogram), (c) image reconstructed from the sinogram (d) corrupted by the loss of 55% (of its randomly selected rows (projections); e) a sinogram recovered from (d) using the iterative band-limited interpolation algorithm and (f) a plot of standard deviation of slice reconstruction error as a function of the iteration number.

7. Applications in Traffic Scenarios

Based on the motion estimation and video enhancement technological achievements, presented in the previous chapters, this chapter describes the utilization of those methods for traffic applications. Related works, till now, have analyzed traffic video streams by independently tracking the different trajectories in the scene. This task is computationally intense, which, in complex scenarios, might not be feasible under real time constrains. The method, unfolded in this chapter, suggests analyzing the entire motion field by means of classical traffic flow parameters extracted by digital image processing means. Those traffic flow parameters are than analyzed for automatic incident detection and quality of service applications. A validation for the image processing method is also provided when the parameters acquired by image processing methods are compared with real-life data.

7.1. System Description and Processing Principles

Figure 7-1 illustrates the processing flow diagram of the traffic video feed. First the video feed is stabilized and its resolution is enhanced as described in Ch. 4. and Ch. 5. respectively. The enhanced output is presented to the operator. Both stabilization and SR processing algorithms extract the real-motion parameters from the processed stream. Those parameters are than translated into traffic stream variables: speed, flow, and concentration. Traffic flow models draw the relationships among the traffic stream variables, and characterize the road and its condition. The road conditions, as they are manifested in the model's parameters, are used for detection of incidents and abnormal behavior. Sect. 7.2 introduces traffic theory and Sect. 7.3 describes how traffic parameters are revealed in data acquired by video-based transportation systems, based on the image processing methods that were presented in the previous chapters. Sect. 7.4 provides a validation of the method, while Sect. 7.5 presents a 3D traffic model which is utilized for estimation of road conditions and for automatic incident detection (AID) applications.

7.2. Characterization of Traffic Streams in Video

Traffic models describe the relationship among traffic stream characteristics. The items of interest in traffic theory have been the following:

- rates of flow (vehicles per unit time);
- speeds (distance per unit time);
- travel time over a known length of road (or something the inverse of speed, "tardity" is used);
- occupancy (percent of time a point on the road is occupied by vehicles);
- density (vehicles per unit distance);
- time headway between vehicles (time per vehicle);
- spacing, or space headway between vehicles (distance per vehicle); and
- concentration (measured by density or occupancy).

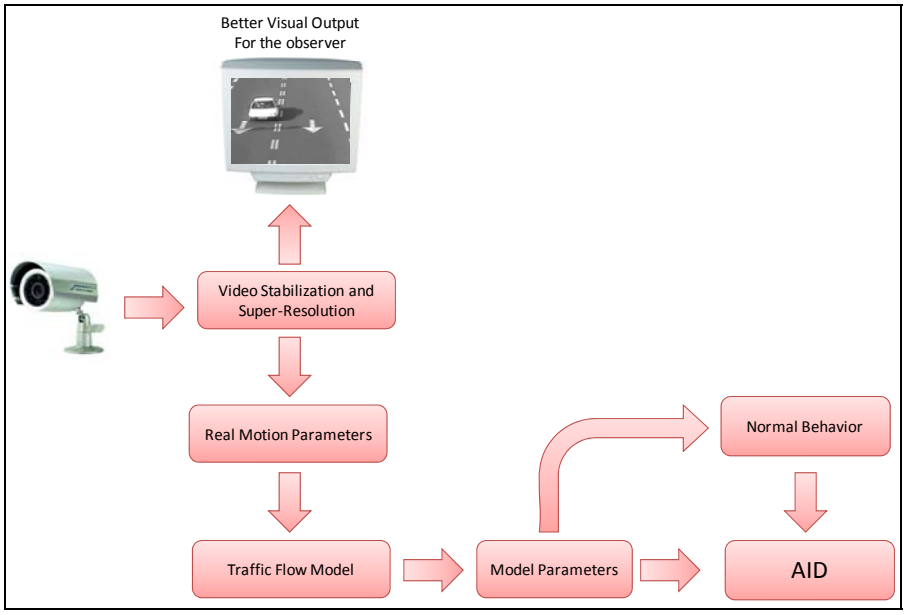


Figure 7-1 – Traffic Video Processing flow diagram

In discussing the models, the link between theory and measurement capability is important since often theory depends on measurement capability. A general notion of these variables, based on the intuitive idea self-evident from their names, will suffice for the purpose of discussing their measurement. Precise definitions of the variables of interest, which are used within the scope of the research, are given in Sect. 7.3. Five measurement procedures are discussed in this section:

- measurement at a point;

- measurement over a short section (less than 10 meters);
- measurement over a length of road (more than 0.5 kilometer (km));
- the use of an observer moving in the traffic stream; and
- wide-area samples obtained simultaneously from a number of vehicles.

The types of measurement are illustrated with respect to a space-time diagram in Figure 7-2. The vertical axis of this diagram represents distance from some arbitrary starting point along the road, in the direction of travel. The horizontal axis represents elapsed time from some arbitrary starting time. Each line within the graph represents the trajectory of an individual vehicle, as it moves down the road over time. The slope of the line is that vehicle's velocity. Where lines cross, a faster vehicle has overtaken and passed a slower one. (The two vehicles do not in fact occupy the same point at the same time.)

Measurement at a point is represented by a horizontal line across the vehicular trajectories: the location is constant, but time varies. In its earliest applications, video cameras were used to acquire the data in the field, which was then subsequently played back in a lab for analysis. In these early implementations, as illustrated in Figure 7-3, lines were drawn on the video monitor screen (literally, when manual data reduction was used). More recently this has been automated, which nowadays allows the data reduction to be conducted simultaneously with the data acquisition [66].

Measurement over a short section is represented by two parallel horizontal lines a very short distance apart. With video camera technology, two detector 'lines' placed close together provide the same capability for measuring speeds.

A vertical line represents measurement along a length of road, at one instant of time, such as a single snapshot taken from above the road. Measurements along a length of road come either from aerial photography, or from cameras mounted on tall buildings or poles. On the basis of a single frame from such sources, only density can be measured. Once several frames are available, speeds can also be measured

The moving observer technique is represented by one of the vehicle trajectories, the heavy line in Figure 7-2. The wide-area samples are similar to having a number of moving observers at various points and times within the system.

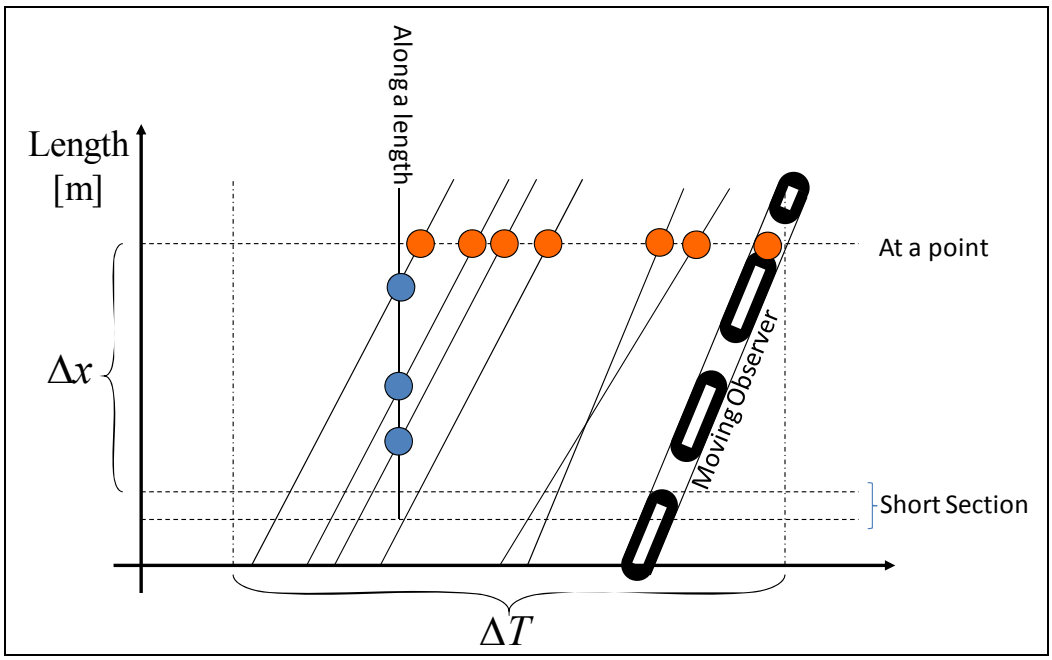


Figure 7-2 – Methods for obtaining traffic data

Figure 7-3 – Traffic measurement at a point in video sequences.

7.3. Variables of Interests

In general, traffic streams are not uniform, but vary over both space and time. Because of that, measurement of the variables of interest for traffic flow theory is in fact the sampling of a random variable. In reality, the traffic characteristics that are labeled as flow, speed, and concentration are parameters of statistical distributions, not absolute numbers.

7.3.1. Flow Rates

Flow rates are collected directly through point measurements, and by definition require measurement over time. They cannot be estimated from a single snapshot of a length of road. Flow rate, q, is the number of vehicles counted, divided by the elapsed time, T:

$$q = \frac{N}{T} \tag{7-1}$$

Flow rates are usually expressed in terms of vehicles per hour, although the actual measurement interval can be much less. Figure 7-4 illustrates a time-space diagram, where the vertical and horizontal axes are the temporal and spatial axes respectively. The diagram depicts, for each time value (frame), the gray-level values of the pixels along the virtual line, Ω, which is presented in Figure 7-3. The number of the cars which passed, through the line throughout the measuring period, T, can be measured by evaluating the temporal cross sections of the diagram. Such a cross-section is presented in Figure 7-5. When evaluating the gray-levels crossings, one can conclude that it might be difficult to extract the number of passing cars from the time-space diagram of the original sequence.

In Sect. 4.3 the formulation of Real-Motion-Separation-Mask (RMSM) was detailed. This tool robustly extracts real-moving objects in real-time. The RMSM of Figure 7-4 is depicted in Figure 7-6, where the line, Ω, is drawn in black and real moving objects are represented in darker pixels. The corresponding RMSM's time-space diagram and its cross-section are presented in Figure 7-7 and Figure 7-8, respectively. While the car counting task is difficult when considering the gray-level values' time-space diagram, (illustrated in Figure 7-5), when considering the RMSM's diagram (Figure 7-8), this task becomes significantly easier. Therefore, it is evident that the RMSM, developed in Sect. 4.3 for stabilization and SR purposes, is a strong tool for flow measurement.

In order to proceed further and simplify the flow computation even more, Athol's assumption [165] of uniform vehicle length is relaxed to a uniform width. Under this assumption, the flow is given by:

$$q = \frac{\sum_{x,y\in\Omega,t\in T}\left[\left(RMSM(x,y,t) - RMSM(x,y,t-1)\right) > Thr\right]}{\alpha T} \tag{7-2}$$

where $\left[\left(RMSM\left(x,y,t\right)-RMSM\left(x,y,t-1\right)\right)>Thr\right]$ is the number of pixels, with change in motion larger than a predefined threshold, along the virtual line, Ω, in time period T and α is the average car's width. Without loss of generality α is set to be 1. Additionally, since frames are acquired in fixed time periods, the elapsed time, T, is measured by the number of frames. By virtue of these assumptions, the flow can be computed from a video sequence by:

$$q=\frac{\sum_{x,y\in\Omega,t\in T}\left[\left(RMSM\left(x,y,t\right)-RMSM\left(x,y,t-1\right)\right)>Thr\right]}{N_{frames}} \tag{7-3}$$

Figure 7-9 illustrates the computation of the flow-rate in two real-life traffic scenarios. Figure 7-9(a) and (b) are frames extracted from real-life video feed of Ayalon highway in free-flow and congestion conditions ([166]). Figures (c) and (d) are the corresponding RMSM frames, where real motion corresponds to brighter pixels. The video frame rate is 5 Hz, hence 300 frames per minute. The flow-rates computed over 300 frames, which figures (a) and (b) where taken from, are shown in figures (e) and (f). The average flow rates for each set are presented in Table 7-1. It is quite evident that the flow-rates for the situation described in figure (b) are higher.

7.3.2. Speeds

Measurement of the speed of an individual vehicle requires observation over both time and space. The instantaneous speed of an individual vehicle is defined as:

$$u_i=\frac{dx}{dt}=\lim_{(t_2-t_1)\to 0}\left(\frac{x(t_2)-x(t_1)}{t_2-t_1}\right) \tag{7-4}$$

In the literature, the distinction has frequently been made between different ways of calculating the average speed of a set of vehicles. The kind of difference that can arise from different methods can be illustrated by the following example. If a traveler goes from A to B, a distance of 20 km, at an average speed of 80 kilometers per hour (km/h), and returns at an average speed of 40 km/h, what is the average speed for the round trip? The answer is of course not 60 km/h; that is the speed that would be found by someone standing at the roadside with a radar gun, catching this car on both directions of the journey, and averaging the two observations. The trip, however, took

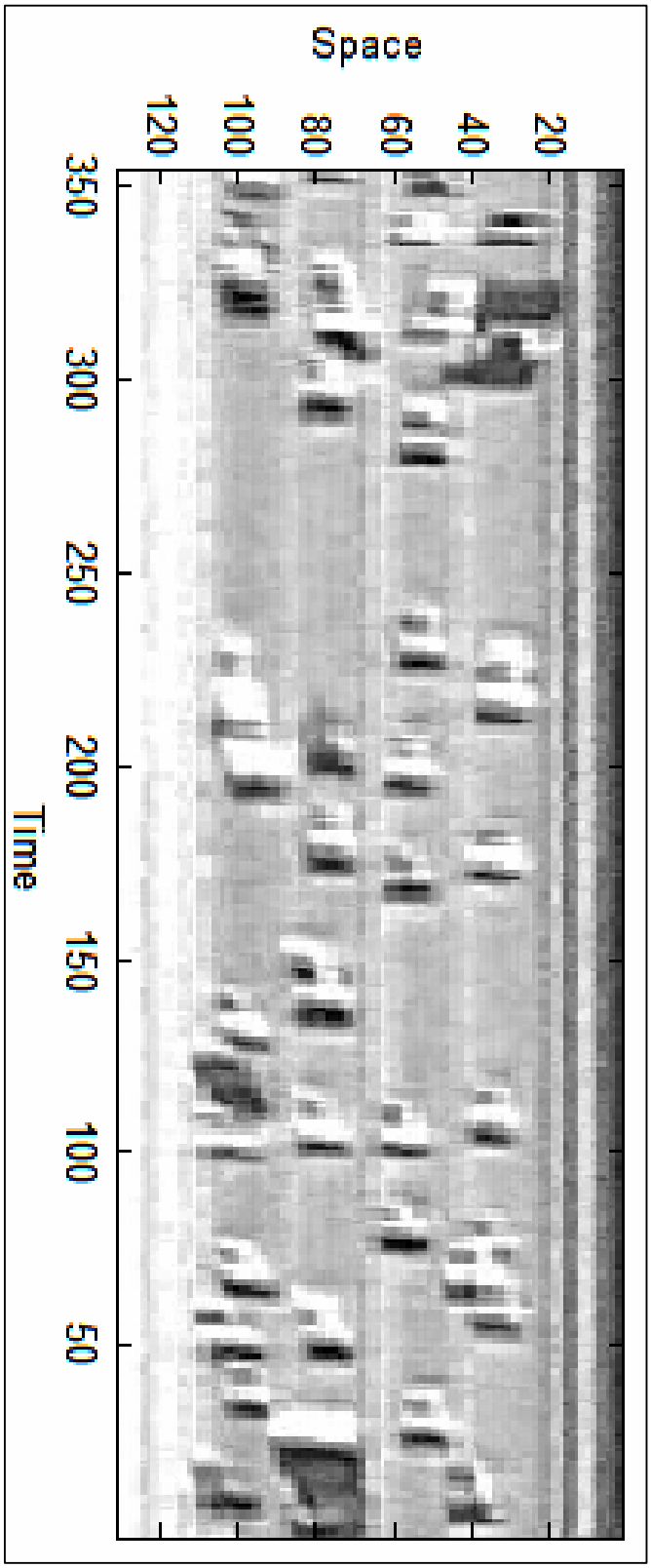


Figure 7-4 – Time – Flow measurement through space-time diagram. The diagram represents the gray-level values of the virtual line presented in Figure 7-3.

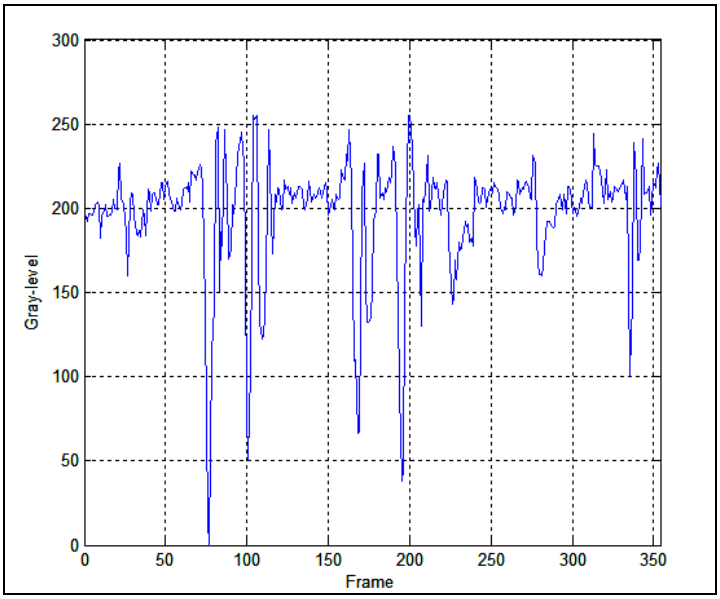


Figure 7-5 – Temporal cross section of the time-space diagram

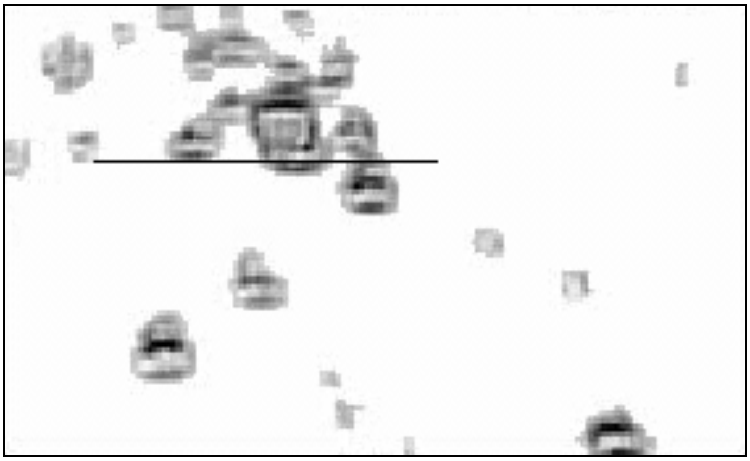


Figure 7-6 – Real Motion Separation Mask (RMSM) of the frame presented in Figure 7-3. Darker pixels represents pixels where motion was detected.

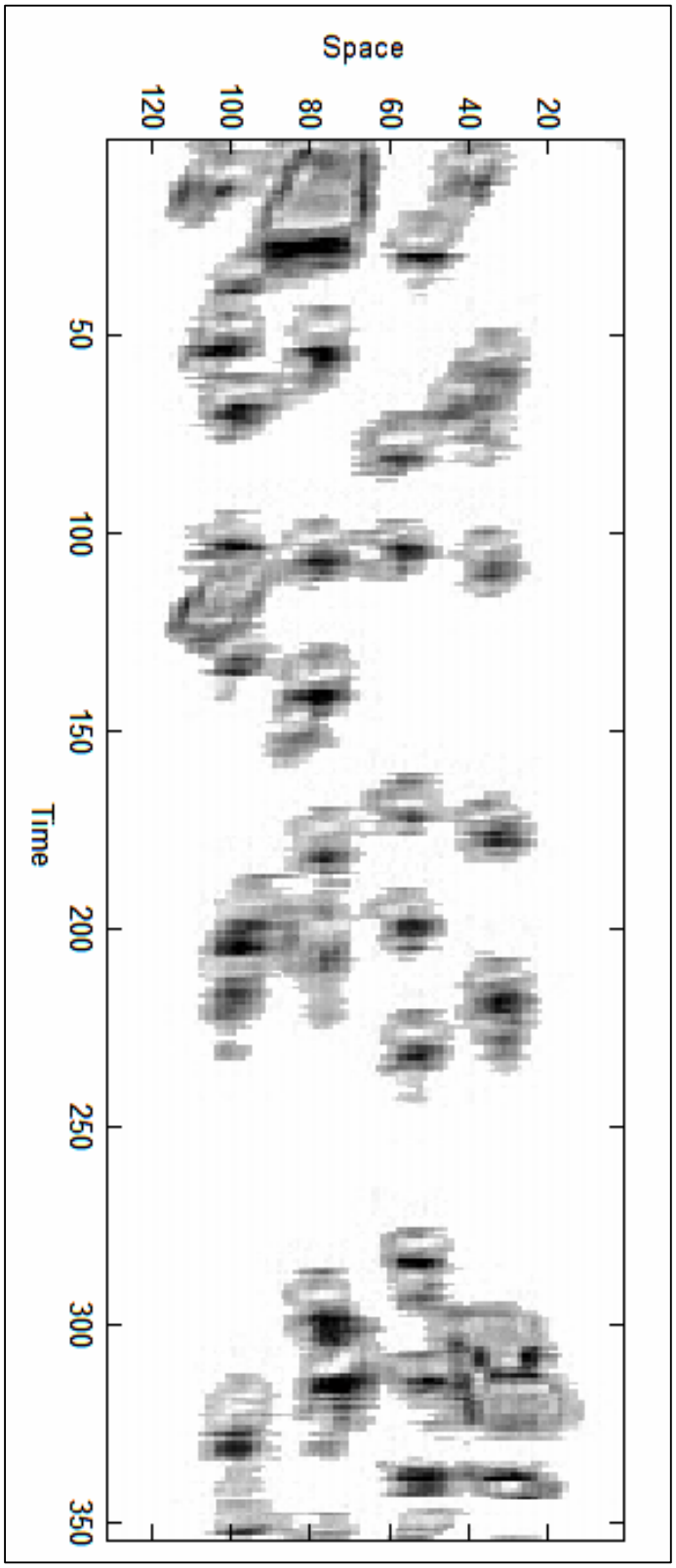


Figure 7-7 - Flow measurement through space-time diagram of the RMSM, presented in Figure 7-6.

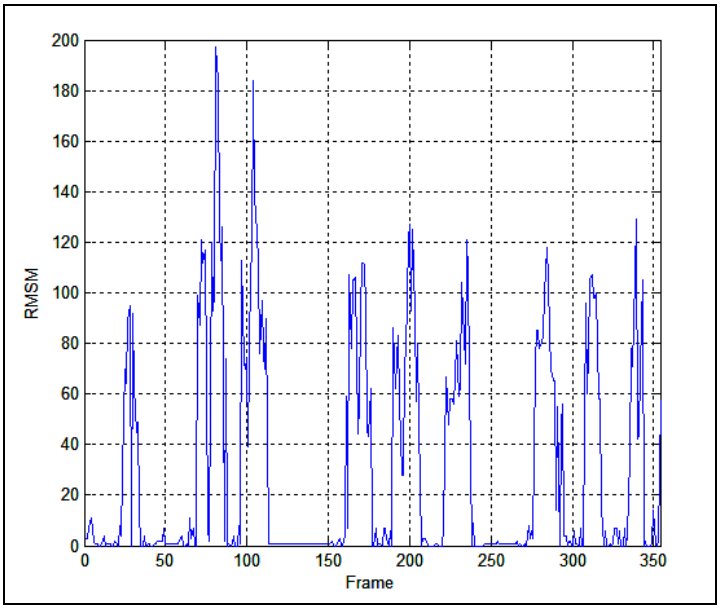


Figure 7-8 - Temporal cross section of the RMSM time-space diagram.

1/4 of an hour one way, and 1/2 an hour for the return, for a total of 3/4 of an hour to go 40 km, resulting in an average speed of 53.3 km/h.

The first way of calculating speeds, namely taking the arithmetic mean of the observation,

$$\overline{u}_t = \frac{1}{N}\sum_{i=1}^{N} u_i \tag{7-5}$$

is termed the time mean speed, because it is an average observations taken over time.

The second term that is used in the literature is space mean speed, but unfortunately there are a variety of definitions for it, not all of which are equivalent. There appear to be two main types of definition. One definition is based on the average time taken to cross a given distance, or space, *D* [167]:

$$\overline{u}_s = \frac{D}{\frac{1}{N}\sum_{i} t_i} \tag{7-6}$$

where t_i is the time for vehicle *I* to cross distance *D*:

$$t_i = \frac{D}{u_i} \tag{7-7}$$

Equation (7-6) is equivalent to using the harmonic mean of the individual vehicle speeds, as follows:

$$\overline{u}_s = \frac{D}{\frac{1}{N}\sum_i t_i} = \frac{D}{\frac{1}{N}\sum_i \frac{D}{u_i}} = \frac{1}{\frac{1}{N}\sum_i \frac{1}{u_i}} \tag{7-8}$$

The second principal type of definition of space mean speed involves taking the average of the speeds of all of the vehicles on a section of road at one instant of time [168]. In Figure 7-2, this method is represented by the vertical line labeled "along a length". In deriving this however, an "isoveloxic" model is assumed, one in which each car follows a linear trajectory in the space time diagram, and is not forced to change speed when overtaking another vehicle. This is equivalent to assuming that the speed distribution does not change with location. A similar definition of space mean speed, without the isoveloxic, is the arithmetic mean of the speeds of vehicles occupying a given length of lane at a given instant [169].

Regardless of the particular definition put forward for mean speed, the individual speeds are computed by the optical flow methods (See Sect. 3.3). The computation of dense optical flow is intensive and might not comply with real-time constrains. To this end, following the discussion above, two methods for extracting the average speed can be utilized. The first method exploits the harmonic mean of speeds measured at a point over time and computes the optical flow only in the vicinity of the virtual line. The second method averages the speeds of all of the vehicles on a section of road at one instant of time, hence one frame. To avoid the situation, where the computation of the optical flow of the entire frame exceeds the system's capability, the two-stage real-motion extraction mechanism, described in 4.3, is utilized and the optical flow is computed only for pixels which are suspected to contain real motion.

Figure 7-10 illustrates the speed rates for 300 frames over two different minutes in which Figure 7-9(a) and (b) were taken in. Evaluating Figure 7-10(a), one can conclude that the traffic conditions described in Figure 7-9(a) allow specific drivers to exceed their speed, while the fluctuations in the graph presented in Figure 7-10(b) suggest that the situation described in Figure 7-9(b) corresponds to "stop and start" [170] conditions, which means that traffic is stopped and released in a periodic manner with varying periods and duty cycles.

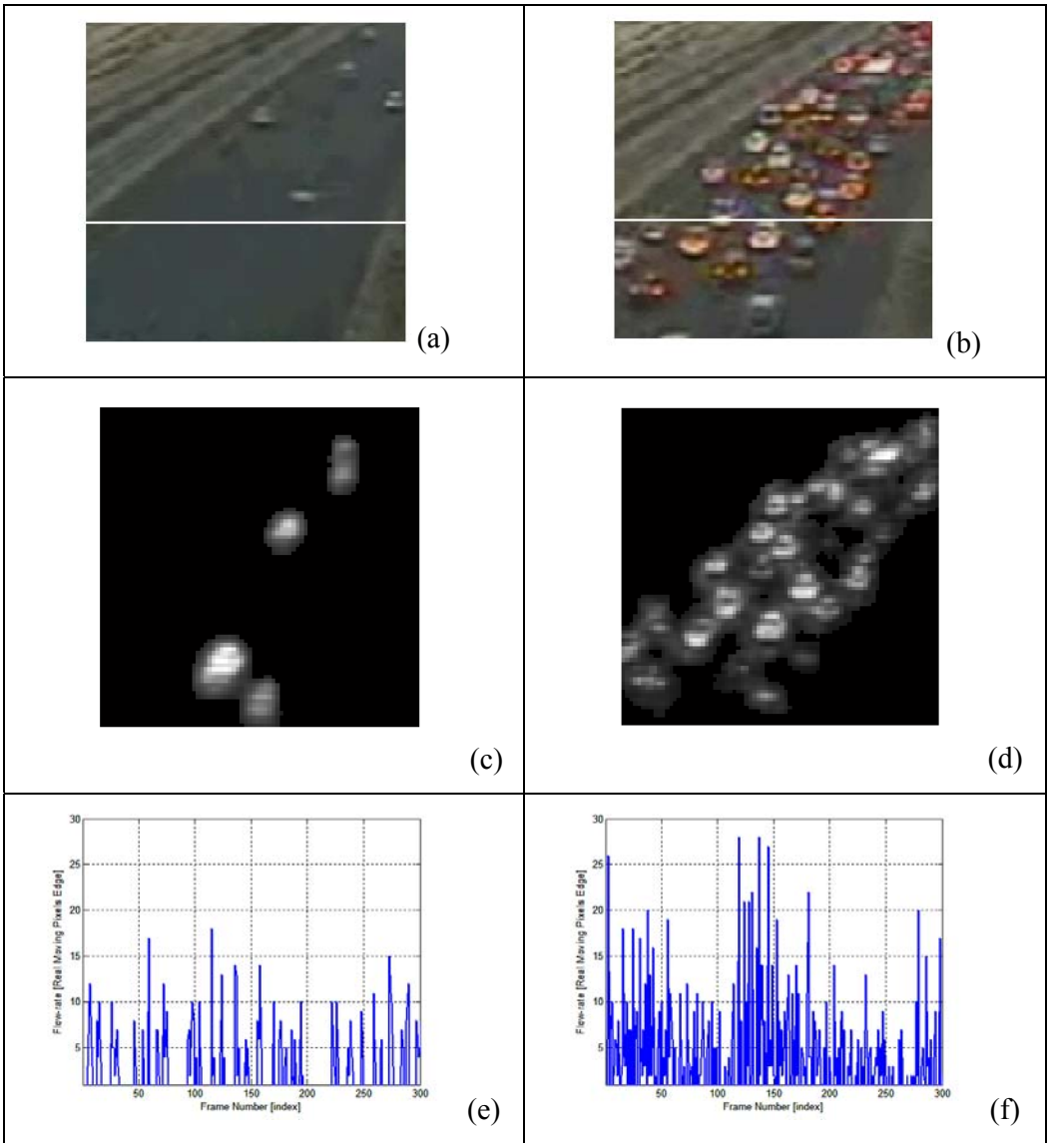


Figure 7-9 – Ayalon highway vision-based flow-rate computation. Figures (a) and (b) are two frames extracted from real-life video feed of Ayalon highway in free-flow and congestion conditions. Figures (c) and (d) are the corresponding RMSM frames. The flow-rates computed over 300 frames of the two different minutes, in which figures (a) and (b) where taken, are shown in figure (e) and (f) respectively.

Table 7-1 – Flow rate computed using a video stream for the two traffic scenarios described in Figure 7-9(a) and (b).

Traffic Condition	Average Flow-rate
Figure 7-9 (a)	2.2458
Figure 7-9 (b)	4.3355

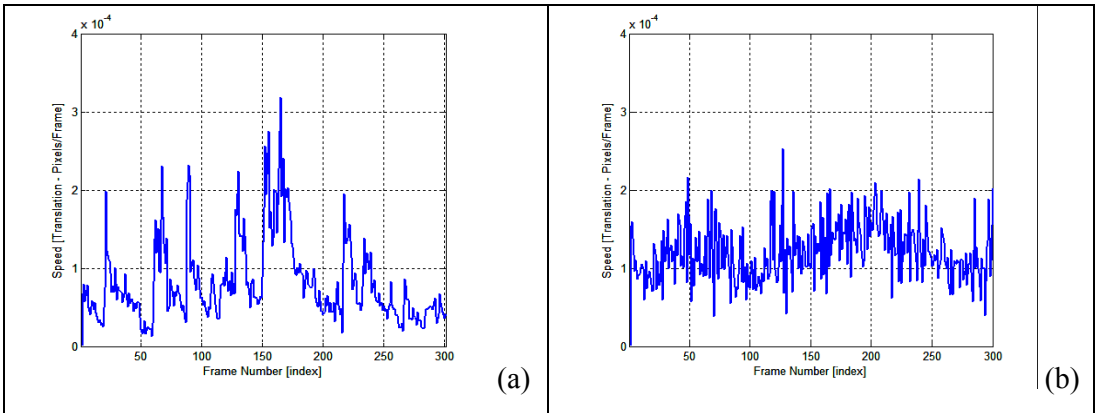


Figure 7-10 – Vision-based Speed rates computed over two different minutes, in which Figure 7-9(a) and (b) were taken in.

7.3.3. Concentration

Concentration has in the past been used as a synonym for density [171]. In this chapter, it seems more useful to use 'concentration' as a broader term encompassing both density and occupancy. The first is a measure of concentration over space; the second measures concentration over time of the same vehicle stream. Density can be *measured* only along a length. If only point measurements are available, density needs to be calculated from speed and flow [171]:

$$k = \frac{q}{\overline{u}_s} \tag{7-9}$$

The difficulty with using this equation for density estimation is that the equation is exactly correct only under some very restricted conditions, or in the limit as both the space and time measurement intervals approach zero. If neither of those situations holds, then the use of the equation to calculate density can give misleading results, which would not agree with empirical measurements. It follows that q equals $u \cdot k$ for the continuous surface, at a point. Real traffic flows, however, are not only made up of finite number of vehicles surrounded by real spaces, but are inherently stochastic [172]. Measured values are averages taken from samples, and are therefore themselves random variables. Measured flows are taken over an interval of time, at a particular place. Measured densities are taken over space at a particular time. Only for stationary processes (in the statistical sense) will the time and space intervals be able to represent conditions at the same point in the time-space plane. Hence it is likely that any measurements that are taken of flow and density (and space mean speed) will not be very good estimates of the expected values that would be defined at the point

of interest in the time space plane – and therefore that Eq. (7-9) will not be consistent with the measured data.

The use of video camera technology allows measuring the actual concentration over a road length in a given time. This is achieved by dividing the number of pixels containing real motion in a given frame by the total number of pixels of the road. Without loss of generality, it is assumed that the road occupies the entire frame. This simplifies the image processing task and eliminates the need to segment the road in the video images, hence concentration is measured by dividing the number of pixels in the real motion separation mask (RMSM), which are equal to 1 by the image size.

Figure 7-11 illustrates the vision-based concentration computation over the 300 frames taken over two different minutes, in which Figure 7-9(a) (blue) and Figure 7-9(b) (red) were taken. The minute average rates are presented in Table 7-2. Evidently the concentration rates in Figure 7-9(b) are higher. Similarly to the notion derived from the speed graphs, the concentration graph also implies that the situation described in Figure 7-9(b) can classified as "stop and start" [170], hence traffic is held and released in a periodic manner with random duty cycles.

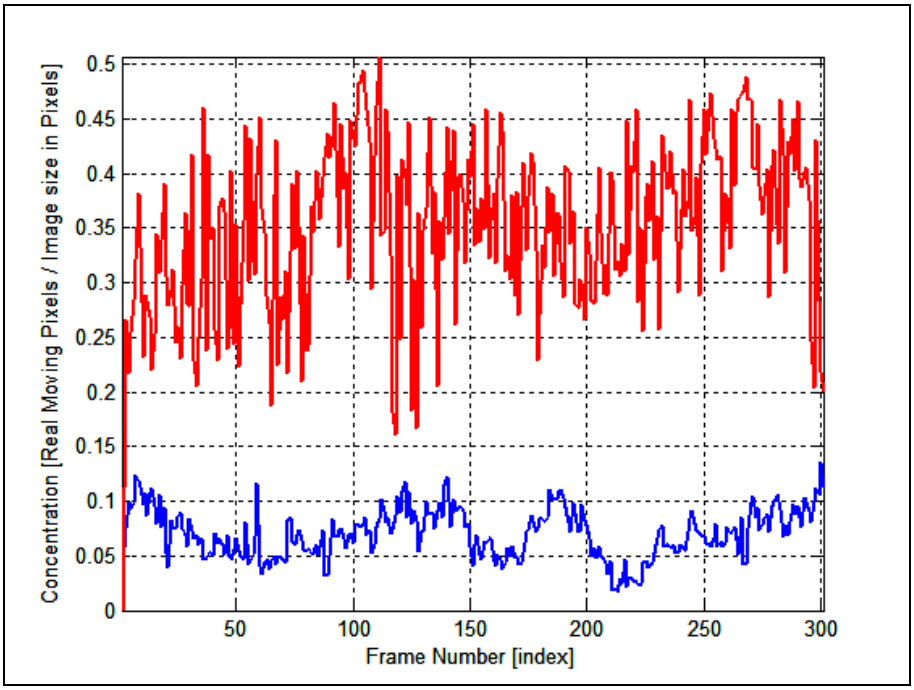


Figure 7-11 – Vision-based Concentration computed over 300 frames taken in the two different minutes in which figures Figure 7-9(a) (blue) and (b) (red) were taken in.

Table 7-2 - Concentration rates computed using a video stream for the two traffic scenarios described in Figure 7-9(a) and (b)

Traffic Condition	Concentration
Figure 7-9(a)	0.0702
Figure 7-9(b)	0.3521

7.4. Demonstration of the Method

In this section the usage of the method, described above, as a tool for traffic parameters estimation, is presented. To this end measurements which were taken in two setups that their parameters are known were used. The first setup is presented in Figure 7-12, while the second is exhibited in Figure 7-13. The setups are located of the highway, so physical measurements can be taken by simple means. In both setup environments the distance between two physical landmarks was physically measured. Vehicles' speeds is calculated by the time it takes each automobile to travel this distance. The camera's frame-rate, in this setup, is 25[Hz], which implies that the vehicles' absolute speeds are given by:

$$\overline{u} = \frac{D}{(T_{leave} - T_{enter})/25}\left[m/Second\right] \tag{7-10}$$

or

$$\overline{u} = \frac{3.6 \cdot D}{(T_{leave} - T_{enter})/25}\left[Km/Hour\right] \tag{7-11}$$

where ***D*** is the measured distance between the two landmarks in each of the setups (5.62[m] and 5.6[m] in the first and second setup scenarios respectively) and T_{enter} is the frame number in which the vehicle passes the first landmark and T_{leave} is the frame number in which it passes the second landmark.

Table 7-3 and Table 7-4 present the *flow, speed* and *concentration* values computed by digital image processing means, as was described in Sect. 7.3. Those were obtained by processing six different segments, one minute long each, extracted from two video sequences which were acquired on the setup locations described in Figure 7-12 and Figure 7-13 (three segments for each location). Flow, in the tables, is given in terms of the total number of pixels that fulfill Eq. (7-3), divided by the total number of frames. The flow, which is 'at the point' measurement, was computed at the line indicates the right landmarks in Figure 7-12 and Figure 7-13. Speed is measured in terms of pixels per frame. The speed, as derived in Sect. 7.3.2, is

computed over the entire time period in terms of the average translation of pixels containing real motion, divided by the number of frames. Finally, concentration, as was described in Sect. 7.3.3, is the average ratio of pixels containing real motion and the entire frame size in pixels. Additionally, the tables contain data acquired by physical measurements (gray cells). The number of cars was physically counted by a human observer and the speed was computed as was described by Eq. (7-11). The speed computed by the image processing methods is presented in terms of the average speed over the entire examined sequence, which in this case, is over 1 minute of the sequence. Therefore the physically measured speed is given both in terms of *average speed per frame*:

$$\tilde{u} = \bar{u}\frac{\#Frames\,containing\,motion}{\#Total\,Frames}\left[Km/Hour\right] \qquad (7\text{-}12)$$

where $\bar{u}$ is the speed computed in Eq. (7-11) and in terms of Km/Hour (in brackets). The time column indicates the minute, of the sequence in which the measurements were taken in. As can be seen, the flow, speed and concentration parameters computed by digital image processing means are highly correlated to the physical measurements. In order to better understand the nature of this correlation a future study of larger scale traffic volumes over a longer period of time has to be carried out. Those future research possibilities are addressed in the discussion.

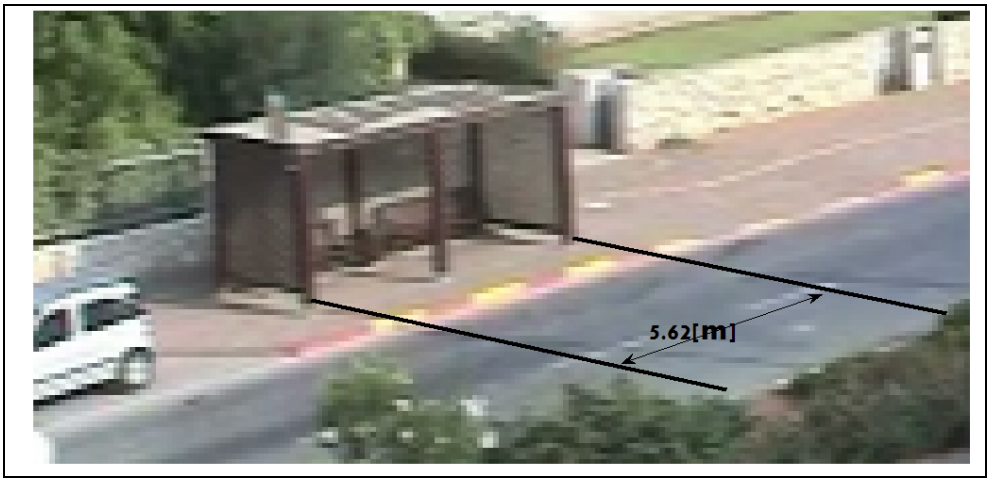


Figure 7-12 – Setup environment 1. The vehicles' speed will be computed by the time it takes a vehicle to pass the marked distance, which was physically measured on site.

7.5. Speed – Flow – Concentration Traffic Model

Since the seminal work of Greenshields [173] a significant amount of work has been invested in effort to establish the relationship between the variables described in the previous section. Some of these efforts begin with mathematical models; others are primarily empirical, with little or no attempt to generalizing.

Figure 7-13 - Setup environment 2.

Table 7-3 – Scenario 1 – Vision based (in white) vs. physical measurements (in gray) computed over 3 different minutes along the video sequence.

Time	Flow[2]	Cars[3]	Speed[4]	Physical Speed[5]	Concentration[6]
2	0.54	9	$1.89 \cdot 10^{-4}$	3.024 (56)	0.0160
4	0.482	8	$1.33 \cdot 10^{-4}$	2.69 (41.56)	0.0162
6	0.458	7	$1.31 \cdot 10^{-4}$	2.35 (45.23)	0.015

Table 7-4 – Scenario 2 – Vision based (in white) vs. physical measurements (in gray) computed over 3 different minutes along the video sequence.

Time	Flow[2]	Cars[3]	Speed[4]	Physical Speed[5]	Concentration[6]
2	0.696	11	$1.49 \cdot 10^{-4}$	3.7 (44)	0.0238
4	1.502	17	$2.6 \cdot 10^{-4}$	5.711 (37.25)	0.0559
6	0.592	10	$1.62 \cdot 10^{-4}$	3.36 (36.8)	0.023

[2] The total number of pixels that fulfill Eq. (7-3), divided by the total number of frames
[3] The total number of counted units.
[4] Frame average translation in pixels.
[5] Km/Hour multiplied by the number of frames with real motion, divided by the number of frames (Km/Hour).
[6] The total number of pixels containing real motion throughout the sequence divided by (Frame size x total number of frames)

The speed-flow relationship is the bivariate relationship on which there has been the greatest amount of work within the past years. The bulk of the recent empirical work on the relationship between speed and flow (as well as the other relationships) was summarized by Hall, Hurdle, and Banks [174]. The authors proposed a model for traffic flow shown in Figure 7-14. The problem for traffic flow theory, in this respect, is that these curves are empirically derived. There is no theory that would explain these particular shapes.

The Speed-concentration model is a linear model, as was stated in the previous section. While there have been studies that claimed to have confirmed this model [175], they tended to have similar sparse portions of the full range of data, usually omitting both the lowest flows and flow in the range near capacity. Additionally, there have also been a number of studies that found contradictory evidence [176].

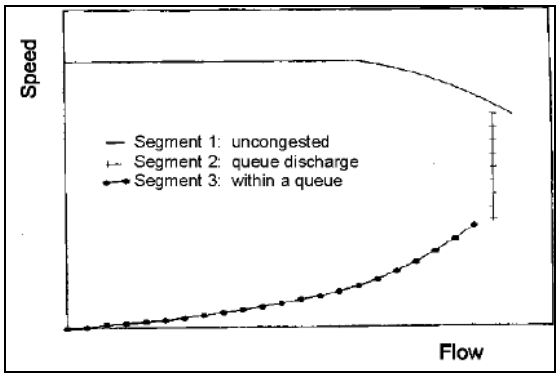


Figure 7-14 – Generalized shape of Speed-flow curve proposed by Hall, Hurdle and Banks ([174])

There are several works that address flow-concentration models [177], all suggesting an empirical inverted V-shape model. The inverted-V model implies that drivers maintain a roughly constant average time gap between their front bumper and the back bumper of the vehicle in front of them provided their speed is less than some critical value. Once their speed reaches this critical value (which is as fast as they want to go), they cease to be sensitive to vehicle spacing.

To conclude, the current status of mathematical models for speed-flow-concentration relationships is in a state of flux. The models that dominated the discourse for nearly 30 years are incompatible with the data currently being obtained, and with currently accepted depictions of speed-flow curves, but no replacement models have yet been developed. Measuring the three parameters, flow, speed and concentration simultaneously, allows indicating the road current conditions over a 3D,

flow-speed-concentration (FSC), space. Recognition of three-dimensional relationships is important for improved understanding. Consequently, it is important to make more use of those sets of freeway data in which all three variables have been measured and no estimation is needed. The simple image processing methods, suggested above, for traffic video analysis offer the ability to do that.

The graph depicted in Figure 7-15 presents 12 points which correspond to the average flow, speed and concentration (FSC) rates computed over a video feed of 12 minutes, where both free-flow and "stop and start" traffic are present. The applicability of the method is illustrated in Figure 7-16 where 3 frames extracted from the video sequence are presented. Figure (a), (b) and (c) are of the frames taken in the corresponding minute, where the FSC parameters for points (A), (B) and (C), shown in Figure 7-15, were computed over. As one can see Figure 7-16(a) presents free-flow road conditions, where vehicles are unimpeded in their ability to maneuver within the traffic stream. Figure 7-16(b), on the other hand, presents road conditions where reasonably free flow and free-flow speeds are maintained, while the concentration rate is relatively higher than the one characterized figure (a). Finally, Figure 7-16(c) presents the road where the concentration rate is relatively high, speeds begin to decline slightly with increasing flows and density begins to increase somewhat more quickly.

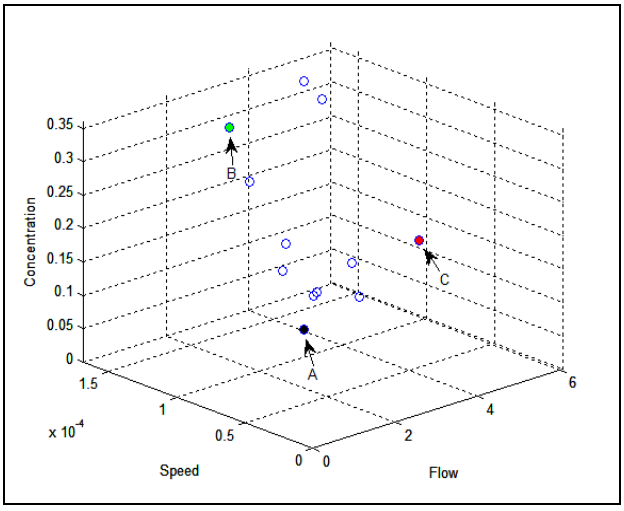
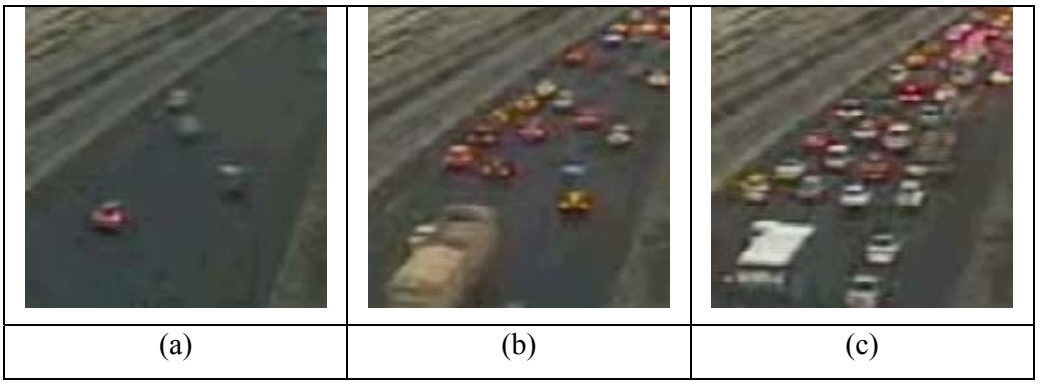


Figure 7-15 - Flow-Speed-Concentration (FSC) 3D model. Each point of the points on the graph represents the floe, speed and concentration average values, computed over a minute is a 12 minutes real-life video feed of Ayalon highway, which contains free-flow as well as "stop and start" traffic. *Speed in given in [Distance in Pixels/Frame], Flow in [Total number of Pixels in the RMSM edge /Frame] and concentration is the average of the ratio (number of Real-moving pixels)/(Frame Size in Pixels).*



(a)	(b)	(c)

Figure 7-16 – The road conditions for the points (A), (B) and (C) in Figure 7-15.

7.6. Traffic Applications

7.6.1. Automatic Incident Detection

Figure 7-17 is a frame extracted from the sequence, one of which earlier frames is given in Figure 7-3. As one can see, an accident is taking place on the right hand side of the image. The effect of the accident on traffic is obvious. For each frame of the sequence the flow, speed and concentration are measured digital image processing means, as described in Sect. 7.2, where flow is computed at the line marked on Figure 7-17. Then, each frame is placed on the FSC-space according to its measured traffic parameters. This is depicted in Figure 7-18. The red markings represent frames taken before the accident took place, while the black ones are of frames taken after the accident occurred. Evaluating Figure 7-18, one can segment the frames projections on the FSC-space. As traffic incidents impact tends to persist for several minutes, automatic incident detection (AID) alarm can be triggered after a certain readings of the parameters are far from the normal, free-flow average, marked with blue circle on the figure.

7.6.2. Quality and Levels of Service

Quality of service requires quantitative measures to characterize operational conditions within a traffic stream. Level of service (LOS) is a quality measure describing operational conditions within a traffic stream, generally in terms of such service measures as speed and travel time, freedom to maneuver, traffic interruptions, and comfort and convenience.

Figure 7-17 – Traffic incident scenario. The accident of the right hand side of the road creates a change in traffic parameter: flow, speed and concentration.

Six LOS are defined for each type of facility that has analysis procedures available [170]. Letters designate each level, from A to F, with LOS A representing the best operating conditions and LOS F the worst. Each level of service represents a range of operating conditions and the driver's perception of those conditions. Figure 7-19, which is taken from [170], visually illustrates the different LOS's.

Each facility type that has a defined method for assessing capacity and level of service also has performance measures that can be calculated. These measures reflect the operating conditions of a facility, given a set of roadway, traffic, and control conditions. Travel speed and density on freeways, delay at signalized intersections, and walking speed for pedestrians are examples of performance measures that characterize flow conditions on a facility. This strongly implies that the levels of service correspond to the facility, or the facility type, examined, and it may vary from one facility to the other. In most cases The LOS are defined to represent reasonable ranges in the three critical flow variables: speed, density, and flow rate. This makes the described image processing methods a great tool for this purpose. The feasibility for that has been demonstrated in Figure 7-15 and Figure 7-16, where the different road conditions, hence class of service, which are depicted in Figure 7-16 are represented in different locations on the 3D model in Figure 7-15.

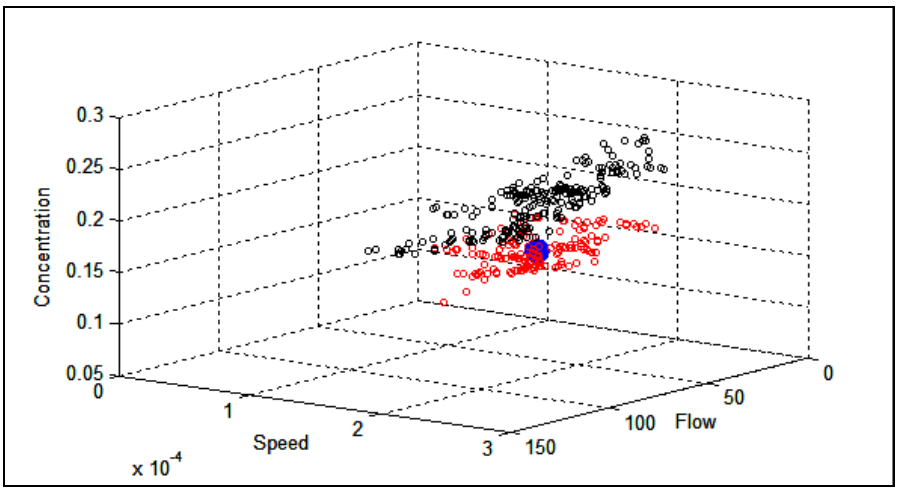


Figure 7-18 – Flow-Speed-Concentration 3D space of a video sequence capturing an accident. The traffic parameters are measured for every frame in the sequence. The red markings are of frames taken before the accident took place, while the black ones are of frames taken after the accident occurred. The blue circle represents the normal situation, free-flow, average of the three traffic parameters. *Speed in given in [Distance in Pixels/Frame], Flow in [Total number of Pixels in the RMSM edge /Frame] and concentration is the average of the ratio (number of Real-moving pixels)/(Frame Size in Pixels).*

7.7. Summary

In this chapter, the basic concepts of traffic theory and their utilization in video based systems are outlined. First, the variables are introduced by which traffic flow is described. Sequentially, the chapter describes how those variables can be measured from traffic video streams. Having the traffic variables measured based on the flow-speed-concentration 3D model, a robust AID scheme is suggested as well as measuring the traffic class of service. The conclusion from the discussion in this chapter is that integration of classical traffic flow models into the decision mechanism of video based intelligent transportation systems shows a great potential for both infrastructure traffic systems, which are sited along the road, and on-platform systems, which are mounted on the car.

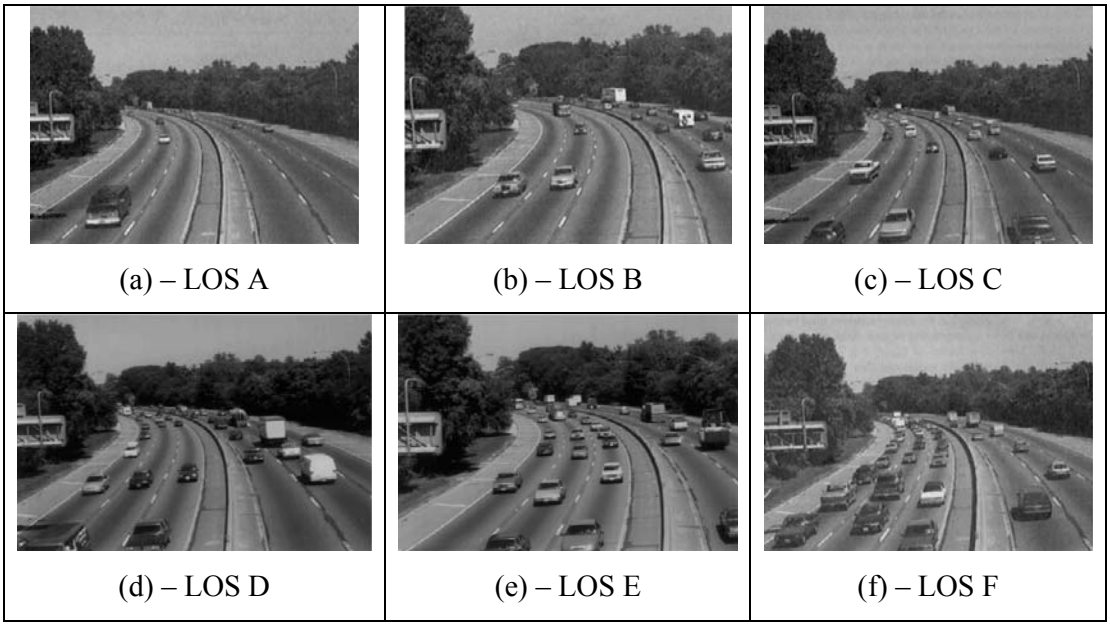


Figure 7-19 – Level of service illustrations, taken from [170].

8. Discussion

This thesis explored the issue of multiframe restoration, enhancement and analysis of videos, acquired in unknown, time-varying system and environmental conditions. Specifically, the presented research addresses the following challenges: (i) efficient and robust motion estimation techniques; (ii) Real time methods for video stabilization and super-resolution (SR) in instable, due to camera and environmental noise, videos; and (iii) Motion-based scene analysis and reasoning for traffic control.

Efficient real motion extraction is a mandatory preliminary requirement in order to secure successful videos restoration, resolution enhancement and scene reasoning. The two classes of motion estimation techniques, block matching and optical flow methods, are evaluated in this respect. To this end, a novel general comparison framework is suggested. Additionally, an improvement, through numerical exact derivation, was suggested for optical flow methods.

Having the motion field in hand, the first objective of this research was a *real-time* video stabilization scheme which preserves real motion in the scene. For preserving real motion in the scene, moving objects are detected and the compensation for the distortion of the turbulence is applied only to the static areas of images. The suggested scheme was evaluated by computer turbulence simulation software that was written for this intent.

Evaluation of a turbulent sequence motion field, allows not only to stabilize it, but to enhance its resolution, which exhibit the second goal of this work. The thesis has shown theoretically and practically that distortion caused by turbulence or camera motions can be used to increase image resolution beyond camera's limitation. Along with the development of real-time methods for image SR, the potential and limitations of utilizing the motion field of instable sequences for SR are sought.

An important part of the SR process is the interpolation method which is used. The interpolation method is important for filling in gaps in given samples. This led to the formulation of the discrete sampling theorem, which is a powerful instrument for signal analysis and reconstruction. It supplies tools for measuring the effective number of samples and iterative, as well as analytical, methods for finding the best interpolation in the RMS sense for any given set of samples and transform domain.

Finally, the technology developed in this research is utilized for traffic video systems. First, simple methods for measuring traffic variables in videos are

introduced. Then, by virtue of basic traffic model, a robust AID and quality of service measurements scheme are presented. The integration of classical traffic flow models into the decision mechanism of video based intelligent transportation systems show a great potential both for infrastructure traffic systems, which are sited on the road, and for on-platform systems, which are mounted on the moving car. For validating the image processing methods as tools for traffic flow parameters estimation a further study should be carried out. In this study the image processing methods should be implemented in a run-time environment and large traffic volumes has to measure both by image processing techniques and other types of sensors, such as loop detectors. The data acquired in the described experiment, both by video and other sensors will then be analyzed to determine the required sampling rate, hence the needed frames-rate for accurate determination of the traffic flow parameters.

In general, the thesis presented novel approaches for multiframe image restoration and enhancement and its utilization for real-time applications. The results presented are of both theoretical and practical interest and offer new efficient tools for substantial improvement of infrastructure of vision-based systems in general and of intelligent transportation systems in particular.

9. Conclusions

Image and video enhancement is a core enabling technology in many fields. This thesis has addressed the fundamental question of multiframe restoration, enhancement and analysis of video streams, acquired in time-varying and unknown system and environmental conditions. The presented research addresses the following challenges: (i) efficient and robust motion estimation techniques; (ii) Real time methods for video stabilization and SR in instable, due to camera and environmental noise, videos; and (iii) Motion-based scene analysis and reasoning.

The thesis presents a novel motion estimation techniques comparison framework and an improvement, through numerical exact derivation, for the optical-flow class of motion estimation techniques. Additionally, by evaluation of the motion field and its statistical analysis, the thesis presents a reliable segmentation of video frames into stable and moving components and subsequently stabilizing images, without harming real moving objects, and improving frames resolution. The potential and limitations of utilizing the motion field of instable sequences for SR are sought as well. An important part in the process of resolution enhancement is signal reconstruction from sparse data accumulated from the set of randomly displaced image frames. The method used is improved by the theory of discrete signal reconstruction from sparse data.

Finally, based on the earlier stages of the research, means for reasoning of the scene observed are developed. This allows detection of irregularity of the motion in the scene. In traffic application, this corresponds to congestion or accidents.

The results present novel approaches and offer new efficient tools for substantial improvement of infrastructure of vision-based systems in general and of intelligent transportation systems in particular.

References

[1] Schultz T. J., "Multiframe blind deconvolution of astronomical images", *J. Opt. Soc. Am. A,* 10:1064-1073, 1993.

[2] Roggemann M. C. and Welsh B., *Imaging Through Turbulence*, CRC Press, Boca Raton, FL, USA, 1996.

[3] Farmer W. M., *The Atmospheric Filter, Volume I - Sources*, JCD publishing, 2001.

[4] Mycek, M-A., Pogue, B. W., *Handbook of biomedical fluorescence* (Marcel Dekker Ed.), New York, Basel, Switerland, 2003.

[5] Levine E. R. and Lueck R. G., "Turbulence measurements from an autonomous underwater vehicle", *J. Atmospheric and Oceanic Technology*, 16:1533-1544, 1999.

[6] Welsh B.M. and Gardner C.S., "Performance analysis of adaptive optics systems using slope sensors", *J. Opt. Soc. Am. A*, 6:1913-1923, 1989.

[7] Ellerbroek B., "First-order performance evaluation of adaptive-optics systems for atmospheric-turbulence compensation in extended-field-of-view astronomical telescopes", *J. Opt. Soc. Am. A*, 11(2):783 ,1994.

[8] Farmer W. M., *The Atmospheric Filter volume II - Sources*, JCD publishing, 2001.

[9] Lloyd-Hart M. and Milton N. M., "Fundamental limits on isoplanatic correction with, multiconjugate adaptive optics", *J. Opt. Soc. Am. A*, 20(10):1949-1957, 2003.

[10] Le-Roux B., Conan J. M., Kulcsar C., Raynaud H. F., Mugnier L. M. and Fusco T., "Optimal control law for classical and multiconjugate adaptive optics", *J. Opt. Soc. Am. A*, 21(7):1261-1276, 2004.

[11] Sheppard D. G., Hunt B. R., Marcellin M. W, "Iterative Multiframe Super-Resolution Algorithms for Atmospheric Turbulence-Degraded Imagery", *J. Opt. Soc. Am. A*, 15(4):972-992, 1998.

[12] Sadot D. and Kopeika N. S., "Imaging through the atmosphere: practical instrumentation-based theory and verification of aerosol modulation transfer function", *J. Opt. Soc. Am. A*, 10(1):172-179, 1993.

[13] Cohen B. , Avrin V. , Belitsky M., and Dinstein I., "Generation of a restored image from a video sequence recorded under turbulence effects", *Optical Engineering*, 36(12):3312-3317, 1997.

[14] M. C. Roggermann and B. Welsh, *Imaging Through Turbulence*, CRC Press, Inc, 1996

[15] Hadar O., Fisher M. and Kopeika N. S., "Image resolution limits resulting from mechanical vibrations, Part III: numerical calculation of modulation transfer function", *Opt. Eng.*, 31:581-589, 1992.

[16] Hadar O., Dror I., and Kopeika N. S., "Image resolution limits resulting from mechanical vibrations, Part IV: Real-Time numerical calculation of Optical Transfer Functions and experimental verifications", *Opt. Eng.*, 33(2):566-578, 2008.

[17] Uomori, K., Morimura, A., Ishii, H., Sakaguchi, T., and Kitamura, Y., "Automatic image stabilizing system by full-digital signal processing", *IEEE Trans. Consumer Electronics*, 36(3):510–519, 1990.

[18] Engelsberg, A., and Schmidt, G., "A comparative review of digital image stabilising algorithms for mobile video communications", *IEEE Trans. Consumer Electronics*, 45(3):591–597, 1999.

[19] Hansen, M., Anandan, P., Dana, K., van der Wal, G., and Burt, P., "Real-time scene stabilization and mosaic construction", *Proc. second IEEE workshop on applications of computer vision*, Sarasota Florida, USA, 54–62, 1994.

[20] Gullu, M. K., and Erturk, S., "Fuzzy image sequence stabilization", *Electronics Letters*, 39(16):1170–1172, 2003.

[21] Fraser D., Thorpe G. and Lambert A., "Atmospheric turbulence visualization with wide-area motion-blur restoration", *J. Opt. Soc. Am. A*, 16(7):1751-1758, 1999.

[22] Scott-Fleming I., Hege K., Clyde D., Fraser D. and Lambert A., "Gradient based optical flow techniques for tracking image motion due to atmospheric turbulence", *Proc. Signal Recovery and Synthesis Symp.*, Albuquerque, NM, USA, 68-70, 2001.

[23] Yaroslavsky L., Fishbain B., Shteinman A., Gepshtein Sh., "Processing and Fusion of Thermal and Video Sequences for Terrestrial Long Range Observation Systems", *Proc. 7th Int'l Conf. Information Fusion*, Stockholm Sweden, 848-855, 2004.

[24] Gepshtein Sh., Shtainman A., Fishbain B. and Yaroslavsky L. P., "Restoration of atmospheric turbulent video containing real motion using elastic image registration", *Proc. European Signal Processing Conference*, Vienna, Austria, 477-480, 2004.

[25] Frakes D. H., Monaco J. W. and Smith M. J. T., "Suppression of atmospheric turbulence in video using an adaptive control grid interpolation approach", *Proc. IEEE Int'l Conf. on Acoustics, Speech, and Signal Process.*, Salt Lake City, Utah, USA, 3:1881-1884, 2001.

[26] Srinivasan S. and Chappella R., "Image Sequence Stabilization, Mosaicking and Superresolution", *Video and Image Processing Handbook* (A.C. Bovic Ed.), ch. 3.13, 259-268, Academic Press, New York, NY, USA, 2000.

[27] Galatsanos N. P., Wernick M. N. and Katsaggelos A. K., "Multi-channel Image Recovery", *Video and Image Processing Handbook* (A.C. Bovic Ed.), ch. 3.7, 161-174, Academic Press, New York, NY, USA, 2000.

[28] Schultz R.R., "Extraction of High-Resolution Frames from Video Sequences", *IEEE, Trans. Image Process.*, 5(6):996-1011, 1996.

[29] Ozkan M. K., Erdem A. T., Sezan M. I., and Tekalp A. M., "Efficient multiframe Wiener restoration of blurred and noisy images', *IEEE Trans. Image Process.*, 1(4):453-476,1992.

[30] Guan L., "Restoration of randomly blurred Images by the Wiener Filter", *IEEE, Trans. Acoustics, Speech and signal Proc.*, 37(4):589-595, 1989.

[31] Adiv G., "Determining 3-D motion and structure from optical flow generated by several moving objects", *IEEE Trans. Pattern Anal. Machine Intell.*, 7:384-401, 1985.

[32] Capel D. and Zisserman A., "Automated mosaicing with super-resolution zoom", *Proc. Computer Vision and Pattern Recognition*, New York, USA, 885-891, 1998.

[33] Irani M. and Peleg S., "Improving resolution by image registration", *Graph. Models Image Process.*, (53):231-239, 1991.

[34] Alam M. S., Bognar J. G., Hardie R. C. and Yasuda B. J., "High-resolution infrared image reconstruction using multiple randomly shifted low-resolution aliased frames", *Proc. SPIE*, 3036:102-112, 1997.

[35] Hardie R.C., Barnard K. J. and Armstrong E. E., "Joint map registration and high resolution image estimation using a sequence of undersampled images", *IEEE Trans. Image Process.*, 6:1261-1633, 1997.

[36] Elad M. and Feuer A., "Recursive Optical Flow Estimation – Adaptive Filtering Approach", *J. Visual Com. and Image Rep.*, 9(2):119-138, 1998.

[37] Goldberg N., Feuer A. and Goodwin, G.C., "Super - Resolution Reconstruction Using Spatio - Temporal Filtering", *J. of Visual Com. and Image Rep.*, 14:508—525, 2003.

[38] Lin Z., and Shum H., "Fundamental limits of reconstruction-based super-resolution algorithms under local translation", *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):83–97, 2004.

[39] Capel D. and Zisserman A.,"Computer vision applied to superresolution", *IEEE Signal Proc. Mag.*, 20(3):75-86, 2003.

[40] Baker S., and Kanade T., "Limits on super-resolution and how to break them", *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(9):1167–1183, 2002.

[41] Farsiu S., Robinson D., Elad M. and Milanfar P., "Fast and Robust Multiframe Super-resolution", *IEEE Trans. Image Process.*, 13(10):1327-1344, 2004.

[42] Farsiu S., Elad M. and P. Milanfar, "Multiframe Demosaicing and Super-Resolution of Color Images", *IEEE Trans. Image Process.*, 15:141-159, 2006.

[43] Li X., "Super-resolution for Synthetic Zooming", *EURASIP – J. Applied Signal Processing*, 2006(1):230-230, 2006.

[44] Ben-Ezra M., Zomet A. and Nayar S.K, "Video Super-Resolution Using Controlled Subpixel Detector Shifts", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(6):977-987, 2005.

[45] Boult T. E. , Chiang M-C and Micheals R. J., "Super-Resolution via Image Warping Theory, Implementation and Evaluation", *Super-Resolution Imaging*, (Chaudhuri S. Ed.), The Springer International Series in Engineering and Computer Science, Springer, 2001.

[46] Chiang, M. and Boult, T., "The integrating resampler and efficient image warping. Proceedings of the DARPA Image Understanding Workshop, Palm Springs California, USA, 843–849, 1996.

[47] Boult, T. and Wolberg, G., "Local image reconstruction and subpixel restoration algorithms, *Graphical Models and Image Processing*, 55(l):63–77, 1993.

[48] Charnotskii M. I., Myakinin V. A. and Zavorotnyy V. U., "Observation of superresolution in nonisoplanatic imaging through turbulence", *J. Opt. Soc. Am. A*, 7(8):1345-1350, 1990.

[49] Lambert A., Fraser D., Jahromi M. R. S. and Hunt B.R., "Super-resolution in image restoration of wide-area images viewed through atmospheric turbulence", *Proc. SPIE*, 4792:35-43, 2002.

[50] Lambert A. and Fraser D., "Super-resolution in imagery arising from observation through an–isoplanatic distortion", *Proc. SPIE*, 5562-8:65-75, 2002.

[51] Shepard D., "A two-dimensional interpolation function for irregularly-spaced data". *Proc. ACM National Conf.*, New-York, NY, USA, 517-524, 1968.

[52] Lodha S.K., Franke R., "Scattered Data Techniques for Surfaces", *Proc. Conf. Scientific Visualization*, Dagstuhl, Germany, 182-222, 1999.

[53] Landau H., "Necessary Density conditions for sampling and interpolation of certain entire functions", *Acta Math.*, 117:37-52, 1967.

[54] Aldroubi A. and Grochenig K., "Non-uniform sampling and reconstruction in Shift-Invariant Spaces", *SIAM Rev.*, 43(4):585–620, 2001.

[55] Marvasti F., *Nonuniform Sampling* (F. Marvasti Ed.), Kluwer Academic / Plenum Publishers, New-York, NY, USA, 2001.

[56] Unser M., "Splines: a perfect fit for signal and image processing", *IEEE Signal Proc. Mag.*, 16:22-38, 1999.

[57]Lee S., Wolberg G. and Shin S. Y., "Scattered Data Interpolation with Multilevel B-Splines", *IEEE Trans. Visualization Comput. Graph.*, 3:228-244, 1997.

[58] Margolis E., Eldar Y. C., "Interpolation with Non-Uniform B-Splines", *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Montreal, Canada, 2:577-580, 2004.

[59] A. Averbuch, R, Coifman, M. Israeli, I. Sidelnikov and Y. Shkolinsky, Irregular Sampling for Multi-dimensional Polar Processing of Integral Transforms, in: Advances in Signal Transforms: Theory and Applications, (J. Astola and L. Yaroslavsky ,eds), Hindawi Publishing Corp., 2007

[60] Averbuch A. Z. and Zheludev V. A., "Wavelet and Frame Transform Originated from Continuous and Discrete Splines", *Advances in Signal transforms: Theory and Applications* (J. Astola, L. Yaroslavsky, Eds.), EURASIP Book Series on Signal Processing and Communications, Hindawi, 2007.

[61] Cheung S.-C., Kamath C., "Robust Techniques for Back-ground Subtraction in Urban Traffic Video", *Proc. SPIE*, 5308:881-892, 2004.

[62] Wang K., Jia X. and Tang Sh., "A survey of vision-based automatic incident detection technology", *Proc. IEEE Vehicular Electronics and Safety*, Xi'an China, 290- 295, 2005.

[63] Harvey B., Champion G. and Deaver R., "Accuracy of traffic monitoring equipment field tests", *Proc. IEEE Conf. Vehicle Navigation and Information System*, Ottawa Ontario, CA, 141-144, 1993.

[64] Kastrinaki V., Zervakis M., Kalaitzakis K., "A survey of Video Processing Techniques for Traffic Applications", *Image and Vision Computing*, 21:359-381, 2003

[65] Shashua A., Gdalyahu Y., Hayun G., "Pedestrian Detection for Driving Assistance Systems: Single-frame Classification and System Level Performance", *Proc. IEEE Intelligent Vehicles Symp.*, Parma Italy, 1-6, 2004.

[66] Tseng B. L., Lin C-Y, Smith J. R., "Real-Time Video Surveillance for Traffic Monitoring Using Virtual Line Analysis", *Proc. IEEE International Multimedia and Expo*, Lausanne Switzerland, 2:541-544, 2002.

[67] Smith S.M. and Brady J.M., "ASSET-2: Real-Time Motion Segmentation and Shape Tracking", *IEEE Trans. Pattern Anal. Machine Intell.*, 17(8):814-820, 1995.

[68] Irani M. and Peleg S., "Image sequence enhancement using multiple motions analysis", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Champaign, IL, USA, 216-221, 1992.

[69] X. Wang and J. Zhang, "A Traffic Detection Method Based on Wavelet Mallat Algorithm", IEEE Mid-Summer Workshop on Soft Computing in Industrial Applications, Helsinki, Finland, 2005.

[70] Gartner N. H., Rathi A.J. and Messer C. J. (Eds.): Revised Monograph on Traffic Flow Theory: A State-of-the-Art Report, *Special Report by the Transportation Research Board of the National Research Council*, 2005. (http://www.tfhrc.gov/its/tft/tft.htm)

[71] Yaroslavsky L. P. and Caulfield H. J., "Deconvolution of multiple images of the same object", *J. Applied Optics*, 33:2157-2162, 1994.

[72] Ozkan M. K., Erdem A. T., Sezan M. I., and Tekalp A. M., "Efficient multiframe Wiener restoration of blurred and noisy images', *IEEE Trans. Image Process.*, 1(4):453-476,1992.

[73] Hunt B. R. and Kubler 0., "Karhunen-Loeve multispectral image restoration. Part I: Theory' *IEEE Trans. Acoustic, Speech, Signal Process.*, 32:592-599, 1984.

[74] Ghiglia D., "Space-invariant deblurring given N independently blurred images of a common object', *J. Opt. Soc. Am. A*, 1:398-402, 1982.

[75] Tsai R. and Huang T., "Multiframe image restoration and registration", *Advance in Computer Vision and Image Processing Vol. 1*, JAI Press, 1984.

[76] Galatsanos N. P., Katsaggelos A. K., Chin R. T., and Hillery A. D., "Least squares restoration of multichannel images", *IEEE Trans. Signal Process.*, 39:2222-2236, 1991.

[77] Li R., Zeng B. and Liou M. L., "A New Three-Step Search Algorithm for Block Motion Estimation", *IEEE Trans. Circuits And Systems Video Tech.*, 4(4):438-442, 1994.

[78] Lu J. and Liou M. L., "A Simple and Efficent Search Algorithm for Block-Matching Motion Estimation", *IEEE Trans. Circuits And Systems Video Tech.*, 7(2):429-433, 1997.

[79] Po L.-M. and Ma W.-C., "A Novel Four-Step Search Algorithm for Fast Block Motion Estimation", *IEEE Transactions Circuits And Systems Video Tech.*, 6(3):313-317, 1996.

[80] Zhu Sh. and Ma K.-K., "A New Diamond Search Algorithm for Fast Block-Matching Motion Estimation", *IEEE Trans. Image Process.*, 9(2):287-290, 2000.

[81] Nie Y. and Ma K.-K., "Adaptive Rood Pattern Search for Fast Block-Matching Motion Estimation", *IEEE Trans. Image Process.*, 11(12):1442-1448, 2002.

[82] Chen Z., Xu J., He Y. and Zheng J., "Fast integer-pel and fractional-pel motion estimation for H.264/AVC.", *J. Vis. Commun. Image Represent.*, 17:264–290, 2006.

[83] Tourapis A. M., Au O. C. and Liou M. L., "Highly efficient predictive zonal algorithm for fast block-matching motion estimation", *IEEE Trans. Circuit and Systems for Video Tech.*, 12:934–947, 2002.

[84] Lucas B.D. and Kanade T., "An Iterative Image Registration Technique with an Application to Stereo Vision", *Proc. the 7th Intl. Joint Conf. on Artificial Intelligence*, Vancouver BC Canada, 674-679, 1981.

[85] Horn B.K.P. and Schunck B., "Determining Optical Flow", *J. Artificial Intelligence*, 17:185-203, 1981.

[86] Amiaz T. and Kiryati N., "Piecewise-smooth dense optical flow via level sets", *Int'l J. Computer Vision*, 68(2):111-124, 2006.

[87] Brox T., Bruhn A., Papenberg N. and Weickert J., "High Accuracy Optical Flow Estimation based on Theory for Wrapping", *Proc. 8th European Conference on Computer Vision*, Prague, Check Republic, 4:25-36, 2004.

[88] Barron L. J., Fleet D. J. and Beachemin S. S., "Performance of Optical Flow Techniques", *Int'l J. Computer Vision*, 12:43-77, 1994.

[89] Barron J. L. and Thacker N. A., "Computing 2D and 3D Optical Flow", *Tutorial Image Science and Biomedical Engineering Division, Medical School, University of Manchester*, 2004 (http://www.tina-vision.net/docs/memos/2004-012.pdf)

[90] Galvin B., McCane B., Novins K., Mason D. and Mills S., "Recovering motion fields: An evaluation of eight optical flow algorithms", *Proc. 9th British Machine Vision*, Southampton UK, 1:195-204, 1998.

[91] McCane B., Novins K., Crannitch D. and Galvin B., "On Benchmarking Optical Flow", *Computer Vision and Image Understanding*, 84:126–143, 2001.

[92] Elad M., *Super-Resolution Reconstruction of Images, PhD Dissertation*, Technion – Israel Institute of Technology, 1996.

[93] Yaroslavsky L.P., Shabat G., Fishbain B., Ideses I.A., "Super-Resolution of Turbulent Video: Potentials and Limitations", *Proc. SPIE*, 6812:5-13, 2008.

[94] Goodman J. W., *Introduction to Fourier Optics (2nd Ed.)*, McGraw-Hill, New-York, 1996.

[95] Tikanov A. and Arsenin V., *Solutions of ill-posed problems*, Winston, Washington D.C., USA, 1977.

[96] Root W. L., "Ill-posedness and precision in object-field reconstruction problems', *J. Opt. Soc. Am. A*, 4:171-179, 1987.

[97] Memin, E. and Perez, P., "Hierarchical estimation and segmentation of dense motion fields" , *Intl. J. of Computer Vision,* 46(2):129–155, 2002.

[98] Alvarez, L., Weickert. J. and Sanchez, J., "Reliable estimation of dense optical flow fields with large displacements", *Intl. J. of Computer Vision*, 39(1):41–56, 2000.

[99] Cohen, I., "Nonlinear variational method for optical flow computation", *Proc. Eighth Scandinavian Conference on Image Analysis*, Norway, 1:523–530, 1993.

[100] Weber, J. and Malik, J., "Robust computation of optical flow in a multi-scale differential framework", *Intl. J. of Computer Vision*, 14(1):67-81,1995.

[101] Black, M. J. and Anandan, P., "The robust estimation of multiple motions: parametric and piecewise smooth flow fields", *Computer Vision and Image Understanding*, 63(1):75–104, 1996.

[102] Otte, M. and Nagel, H., Estimation of optical flow based on higher-order spatiotemporal derivatives in interlaced and non-interlaced image sequences, *Artificial Intelligence*, 78:5-43, 1995.

[103] Brox T., Bruhn A. and Weickert J., "Variational motion segmentation with level sets", Proc. European Conference on Computer Vision, Graz Austria, 471-483, 2006.

[104] Bruhn, A. and Weickert, J., "Towards Ultimate Motion Estimation: Combining Highest Accuracy with Real-Time Performance", *Proc. IEEE Intl. Conf. Computer Vision*, 1:749-755, 2005.

[105] Fleet, D.J. and Jepson, A.D., "Computation of component image velocity from local phase information", *Intl. J. Computer Vision*, 5(1):77–104, 1990.

[106] Heeger, D., "Model for the Extraction of Image Flow", *J. Optical Soc. Am. A*, 4(8):1455-1471, 1987.

[107] Nagel, H. H., "Displacement vectors derived from second order intensity variations in image sequences", *Computer Vision, Graphics and Image Processing*, 21:85-117, 1983.

[108] Uras, S., Girosi, F., Verri, A. and Torre, V., "A computational approach to motion perception", *Biological Cybernetics*, 60:79-97, 1988.

[109] Anandan, P., "A computational framework and an algorithm for the measurement of visual Motion", *Intl. J. Computer Vision*, 2:283–310, 1989.

[110] Singh, A., "An estimation-theoretic framework for image flow computation", *Proc. IEEE Intl. Conf. Computer Vision*, Osaka, Japan, 168-177, 1990.

[111] Waxman, A.M., Wu, J. and Bergholm, F., "Convected activation profiles and the measurement of visual motion", *Proc. IEEE Intl. Conf. Computer Vision*, 717-723, 1988.

[112] Bober, M. and Kittler, J., "Robust Motion Analysis", *Proc. IEEE Intl. Conf. Computer Vision*, Seattle, Washington, USA, 947-952, 1994.

[113] Liu, H.C, Hong, T.H., Herman, M. and Chellappa, R., "A General Motion Model and Spatiotemporal Filters for Computing Optical-Flow", *Intl. J. Computer Vision*, 22(2):141-172, 1997.

[114] Camus, T., "Real-Time Quantized Optical Flow", *Proc. IEEE Conf. Computer Architectures for Machine Perception*, 126-131, 1995.

[115] Liu, H., Hong, T.H., Herman, M., Camus, T. and Chellappa, R., "Accuracy vs Efficiency Trade-offs in Optical Flow Algorithms", *Computer Vision and Image Understanding*, 72(3):271-286, 1996.

[116] Simoncelli E.P., "Design of multi-dimensional derivative filters", *Proc. IEEE Int'l. Conf. Image Process.*, Austin, TX, USA, 1:790-793, 1994.

[117] Yaroslavsky L.P., Agranovich A., Fishbain B., Ideses I. A., "Precise differentiation can significantly improve the accuracy of optical flow measurements", *Proc. SPIE*, 6812:11-19 ,2008.

[118] Yaroslavsky L., "Fast Discrete Sinc-Interpolation: A Gold Standard for Image Resampling", *Advances in Signal transforms: Theory and Applications* (Astola J. and Yaroslavsky L. Eds.), EURASIP Book Series on Signal Processing and Communications, Hindawi, 2007.

[119] Elad, M., Teo, P. and Hel-Or, Y., "On the Design of Filters for Gradient-Based Motion Estimation", *Intl. J. Mathematical Imaging and Vision*, 23:345–365, 2005.

[120] Brandt, J. W., "Improved accuracy in gradient-based optical flow estimation", *Intl. J. Computer Vision*, 25(1), 5-22 (1997)

[121] Kearney, J.K., Thompson, W.B. and Boley, D.L., "Optical flow estimation: An error analysis of gradient-based methods with local optimization", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(2):229–244, 1987.

[122] Press W.H., Flannery B.P., Teukolsky S.A., Vetterling W.T., *Numerical recipes. The art of scientific computing*, Cambridge University Press, Cambridge, 1987.

[123] http://www.eng.tau.ac.il/~barak/RealTimeTurbulenceCompensation

[124] Yaroslavsky L.P., Fishbain B., Ideses I., Slasky D., Hadas Z., "Simple methods for real time stabilization of turbulent video", *Proc. ICO Topical Meeting on Optoinformatics*, Saint-Petersburg Russia, 138-140, 2006.

[125] Fishbain B., Yaroslavsky L. P., Ideses I. A., Shtern A., Ben-Zvi O., "Real-time stabilization of long-range observation system turbulent video", *Proc. SPIE*, 6496:60C-71C, 2007.

[126] Fishbain B., Yaroslavsky L. P. and Ideses I. A., "Real-time stabilization of long range observation system turbulent video", *J. Real-Time Image Processing*, 2(1):11-22, 2007.

[127] Fishbain B., Yaroslavsky L.P. and Ideses I.A., "Spatial, Temporal, and Interchannel Image Data Fusion for Long-Distance Terrestrial Observation Systems", *Advances in Optical Technologies*, 2008:Article ID 546808, 2008.

[128] Winkler S., *Digital Video Quality - Vision Models and Metrics*, John Wiley & Sons, 2005.

[129] Huang T. S., Yang G. J. and Yang G. Y., "A Fast Two-Dimensional Median Filtering Algorithm", *IEEE Trans. Acoust. Speech Signal Process.*, 27:13-18, 1979.

[130] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression Video Coding for Next-generation Multimedia*, John Wiley, England 2003.

[131] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards*, Norwell, MA: Kluwer, 1995.

[132] Charnotskii M. I., Myakinin V. A. and Zavorotnyy V. U., "Observation of superresolution in nonisoplanatic imaging through turbulence", *J. Opt. Soc. Am. A*, 7(8):1345-1350, 1990.

[133] Fishbain B., Yaroslavsky L.P., Ideses I.A., "Real-time Turbulent Video Perfecting by Image Stabilization and Super-resolution", *Proc. 7th IASTED Int'l Conf. Visualization, Imaging and Image Processing*, Palma de Mallorca Spain, 213-218, 2007.

[134] Yaroslavsky L.P., Fishbain B., Shabat G. and Ideses I., "Super-resolution in turbulent videos: making profit from damage", *Optics Letters*, 32(20):3038-3040, 2007.

[135] Ferreira P.J.S.G., "Interpolation and the Discrete Papoulis-Gerchberg Algorithm", *IEEE Trans. Signal Process.*, 42(10):2596-2606, 1994.

[136] Papoulis A., "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation", *IEEE Trans. Circuits and Systems*, 22(9):735-742, 1975.

[137] Eskicioglu A. M. and Fisher P. S., "Image Quality Measures and Their Performance", *IEEE Trans. Comm.*, 43(12):2959-2965, 1995.

[138] Pappas T. N. and Safranek R. J., "Perceptual criteria for image quality evaluation", *Handbook of Image and Video Processing* (A. C. Bovik, Ed.), Academic Press, New York, 669-687, 2000.

[139] Lubin J., "The use of psychophysical data and models in the analysis of display system performance", *Digital Images and Human Vision* (A.B. Watson Ed.), 163-178, MIT, Cambridge, MA, 1993.

[140] Li X., "Blind image quality assessment", *Proc. IEEE Intl. Conf. Image Processing* , I449-I452, Rochester, New-York, USA, 2002.

[141] Crete F., Dolmiere T., Ladret P. and Nicolas M., "The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric", *Proc. SPIE*, 6492:0I-17C, 2007.

[142] http://www.eng.tau.ac.il/~barak/EBW/

[143] Fishbain B., Yaroslavsky L. P., Ideses I. A. , Roffet-Crete F., "No-Reference Method for Image Effective Bandwidth Estimation", *Proc. SPIE*, 6808, 2008.

[144] Hadar O. and Boreman G. D., "Oversampling requirements for pixilated-imager systems", *Opt. Eng.*, 38(5):782-785, 1999.

[145] Wharton W. and Howorth D., *Principles of Television Reception*, Pitman Publishing, 1971.

[146] Hamilton E., *JPEG File Interchange Format, Version 1.02*, 1992.

[147] G.M. Callico, A. Nunez, R. P. Llopis, R. Sethuraman, "Low-cost and real-time super-resolution over a video encoder IP", Proceedings. Fourth International Symposium on Quality Electronic Design, pp. 79-84, San-Jose CA-USA, 24-26 March 2003.

[148] Fishbain B., Yaroslavsky L.P., Ideses I. A., "Real Time Turbulent Video Super-Resolution Using MPEG 4", *Proc. SPIE*, 6811:6-9 , 2008.

[149] Mitchell J. L., Pennebaker W. B., Fogg C. E. and LeGall D. J., *MPEG Video Compression Standard*. Chapman and Hall, New York, NY, 1996.

[150] ISO/IEC 14496-2, *Amendment 1, Information technology – coding of audio-visual objects – Part 2:Visual*, 2001.

[151] Ferreira, P. J. S. G., "Iterative and Noniterative Recovery of Missing Samples for 1-D Band-limited Signals", *Nonuniform Sampling* (F. Marvasti Ed.), Kluwer Academic / Plenum Publishers, New-York, NY, USA, 2001.

[152] Gersho A. and Gray R., *Vector Quantization and Signal Compression Boston*, MA: Kluwer, 1992

[153] Compressed sensing resources: http://www.dsp.ece.rice.edu/cs/

[154] Horn R. A. and Johnson C. R., *Topics in matrix analysis*, Cambridge University Press, 1991.

[155] Yaroslavsky L. P., *Digital Holography and Digital Signal Processing*, Kluwer Academic Publishers, Boston, 2004

[156] Gröchenig K. and Strohmer T., "Numerical and Theoretical Aspects of Nonuniform Sampling of Band-Limited Images", *Nonuniform Sampling (F. Marvasti Ed.)*, Kluwer Academic / Plenum Publishers, New-York, NY, USA, 2001.

[157] Kelley B. T. and Madisetti V. K., "The Discrete Radon Transform: Part I – Theory", *IEEE Trans. Image Process.*, 2:382-400, 1993.

[158] Brady M. L., "A fast discrete approximation algorithm for the Radon transform", *SIAM J. Comput.*, 27(1):107–119, 1998.

[159] Brandt A. and Dym J., Fast Calculation of Multiple Line Integrals. *SIAM J. Sci. Comput.*, 20:1417-1429.

[160] Brandt A., Mann J., Brodski M., Galun M., A Fast and Accurate Multilevel Inversion of the Radon Transform. *SIAM J. Appl. Math*, 60(2):437–462, 2000.

[161] Götze W .A. and Druckmüller, H. J., "A fast digital Radon transform – an efficient means for evaluating the Hough Transform", *Pattern Recognition*, 28(12):1985-1992, 1995.

[162] Averbuch A., Coifman R., Donoho D., Israeli M. and Walden J., "Fast Slant Stack: A notion of Radon Transform for data on a Cartesian grid which is Rapidly Computable,Algebraically Exact, Geometrically Faithful, and Invertible", *SIAM Scientific Computing*, 2001.

[163] Averbuch A., Coifman R., Donoho D., Israeli M. and Shkolinsky Y., "A framework for discrete integral transformations I – The pseudo-polar Fourier transform", *SIAM J. Sci. Comput.*, 30(2):764-784.

[164] Standford University, Statistics Department, David Donoho's homepage, http://www.stat.stanford.edu/~donoho/

[165] Athol P., "Interdependance of Certain Operational Characteristics within a Moving Traffic Stream", *Highway Research Record*, 72:58-87, 1965.

[166] Ayalon Highway Authority website: http://www.ayalonhw.co.il

[167] Lighthill M. J. and Whitham G. B., "On Kinematic Waves: Ii. A Theory of Traffic Flow on Long Crowded Roads", *Proceedings of the Royal Society A229 No. 1178*, 317-145, 1955.

[168] Haight F. A., *Mathematical Theories of Traffic Flow*, Academic Press, New-York, 1963.

[169] Gerlough, D. L. and Capelle D. G. (Eds.), *An Introduction to Traffic Flow Theory*, Special Report 79, Highway Research Board, Washington, D.C, 1964.

[170] Dowling R., Rouphail N., Strong D., Schoen J., Troutbeck R. and Bonneson J., *Highway Capacity Manual 2000*, Traffic Research Board (TRB) Publications, 2000

[171] Gerlough D. L. and Huber M. J., *Traffic Flow Theory: a Monograph*, Special Report 165, Highway Research Board, Washington, D.C, 1975.

[172] Newell G. F., *Applications of Queuing Theory*, Second Edition, Chapman and Hall, London, 1982.

[173] Greenshields B. D., "A study of traffic capacity", *Proc. of Highway Research Board 14*, pp. 448-477, 1935.

[174] Hall F. L., Hurdle V. F. and Banks J. H., "Synthesis of Recent Work on the Nature of Speed-Flow and Flow-Occupancy (Or Density) Relationships on Freeways", *Transportation Research Record*, 1365:12-18, TRB, National Research Council, Washington, DC., 1992.

[175] Huber M. J., "Effect of Temporary Bridge on Parkway Performance", *Highway Research Board Bulletin*, 167:73-74, 1957.

[176] Drake J. S., Schofer J. L. and May A. D. "A Statistical Analysis of Speed Density Hypotheses", *Highway Research Record*, 154:53-87, 1967.

[177] Hall F. L. and Gunter M. A., "Further Analysis of the Flow-Concentration Relationship", *Transportation Research Record*, 1091:1-9, 1986.