IMPROVED STEREO VISION ALGORITHMS FOR ROBOT NAVIGATION

by

ALI RANJBARAN

Thesis submitted in fulfillment of the requirements
for the degree of
Doctor of Philosophy

July 2013

ACKNOWLEDGEMENTS

This work could not be done without my supervisor's support, and patience and tolerance of my family.

I would like to thank my supervisor, Dr. Anwar Hasni Abu Hassan for his incredible guidance and support. Without his valuable comments, suggestions and encouragements, this project could not be completed.

I would like to thank my family for their patience and help.

And I would like to express my gratitude to the School of Electrical and Electronic Engineering for giving me this opportunity to carry out my study and research.

TABLE OF CONTENTS

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

LIST OF ABBREVIATIONS

1D	One Dimensional
2D	Two Dimensional
3D	Three Dimensional
AW	Adaptive Weighting
CV	Covariance-Variance
DC	Direct Current
FN	Fuzzy Logic and Neural Network Approaches
NIDAQ	National Instrument Data Acquisition Device
NSCP	Normalized Sum of Cross Products
NSSD	Normalized of Squared Differences
OP	Operational Amplifier
PWM	Pulse Width Modulation
ROF	Rudin, Osher, Fatemi
SAD	Sum of Absolute Differences
SD	Statistical Distance
SSD	Sum of Squared Differences
SVD	Singular Value Decomposition
TV	Total variation
USB	Universal Serial Bus

LIST OF SYMBOLS

$\boldsymbol{a}$	Vector
$A(.)$	Transfer Function
A	Normalizing factor
a	Slop
$a1$	First component of vector $\boldsymbol{a}$
$a2$	Second component of vector $\boldsymbol{a}$
B	Viscous friction
B_L	Blurring Matrix
b_L	Blurring Image
b	Linear Coefficient
C	Convolution Matrix
$b(.)$	Blurred Image
$c1$	First component of Convolution Matrix C
$c2$	Second component of Convolution Matrix C
$c(.)$	Convolution Function
$D(.)$	Motor dead zone
e	Error value
$E(.)$	Error Functions
$f(.)$	Feature Function
$g(.)$	Nonlinear Function
G	Gain value
$G(u)$	Average of image
$h(u)$	Edge detector function
$H(s)$	Transfer function
$Hev(.)$	Heaviside function
$Hist(.)$	Histogram function

I	Intensity value
I_L^{Wr}	Intensity value for the window Wr in left image
I_R^{Ws}	Intensity value for the window Ws in right image
$J1$	SSD Energy Function
$J2$	Gradient Based Energy Function
$J3$	SAD Energy Function
$J4$	Linear Based Energy Function
J	Moment of inertia
$k(.)$	Nonlinear Function
K	Deconvolution Matrix
k_d	Deconvolution parameter
k_t	Torque constant
k_m	Motor constant
$m1$	Minimum point of $J1$ (SSD)
$m2$	Minimum point of $J2$ (Gradient Based)
$m3$	Minimum point of $J3$ (SAD)
$m4$	Minimum point of $J4$ (Linear Based)
M_p	Maximum peak over shout
mot	Mean value
N	Number of pixels
P	Input image
$Q(.)$	Weighting function
q	Output image
R	Resistor
$S(h)$	Separator Function
t	Intensity Values
T	Matrix
T_i	Input torque
T_o	Output torque
T_p	Peak time
T_s	Settling time

thr	Threshold
U^+	Intensity Values
U^-	Intensity Values
u	Image
u_t	Image at time t
U	Matrix
$v(.)$	Convolution Function
V	Covariance Matrix
v	Intensity Values
V_{BE}	Base emitter voltage
V_{in}	Input voltage
V_{out}	Output voltage
W	Image window
$\|W\|$	Size of the window W
Wr	Reference window in left image
Ws	Search window in right image
$X(.)$	Gearbox backlash
x	Coordination
X	Vector
Y	Vector
y	Coordination
$z(.)$	Hidden intensity value
α	Control parameter
β	Control parameter
η	Control parameter
ΩN	Image domain size
μ	Mean value
σ	Variance
∇u	Gradient of image u
$\nabla_x u$	Gradient of image u in x direction
$\nabla_y u$	Gradient of image u in y direction

Ω	Image domain
ξ	Damping factor
ω_n	Natural frequency
ω_d	Resonance frequency
θ_{in}	Input angle
θ_{out}	Output angle

PENAMBAHBAIKAN ALGORITHMA PENGLIHATAN STEREO UNTUK NAVIGASI ROBOT

ABSTRAK

Motivasi utama di dalam penyelidikan ini ialah mencari halatuju terbaik untuk navigasi sebuah robot yang menggunakan penglihatan stereo dengan menyelesaikan masalah dalam mencari nilai perbezaan untuk imej yang mengandungi maklumat yang kurang, imej berhingar dan senget. Kaedah tetingkap penyesuaian dilaksanakan dalam tiga pendekatan. Pendekatan 1 dan 2 menggunakan fungsi kos yang biasa seperti SSD, SAD dan GB (kecerunan). Pendekatan 3 menggunakan fungsi berasaskan linear. Dengan menggunakan kaedah penyesuaian, ralat adalah 12% untuk SSD, 10% untuk fungsi asas-Gradien dan 7% untuk fungsi asas-Linear. SSD adalah 8 kali lebih pantas daripada kaedah asas-Gradien dan 2 kali lebih pantas daripada kaedah asas-Linear. Teknik asas-Linear adalah kaedah yang sesuai untuk aplikasi penglihatan stereo kerana ia adalah 50% lebih tepat daripada SSD. Teknik-teknik penghapusan hingar yang dicadangkan telah dibandingkan dengan kaedah ROF dengan mengira jumlah nilai perubahan untuk imej Lena dan Barbara. Apabila Jumlah Nilai Perubahan (TV) adalah 0.88 untuk imej Lena berhingar, ROF mengurangkan TV kepada 0.21, dan kaedah asas-Linear mengurangkan TV kepada 0.24 dan kaedah asas-Gradien mengurangkan TV kepada 0.26. Apabila TV adalah 0.88 untuk imej Barbara berhingar, ROF mengurangkan TV kepada 0.68, kaedah asas-Linear mengurangkan TV kepada 0.62 dan kaedah asas-Gradien mengurangkan TV kepada 0.65. Sistem pelantar yang stabil untuk menstabilkan garis penglihatan mengufuk telah direkabentuk dan dibangunkan. Sistem ini dilengkapi dengan penstabil miring dan mendatar gelung tertutup menggunakan penderia pecutan dwi-paksi. Ia menstabilkan sudut miring dan mendatar dengan pemalar masa 0.5 saat dan ralat keadaan mantap adalah kurang daripada 0.025 radian. Kualiti anggaran kedalaman dan arah tuju untuk navigasi telah diperbaiki kepada 88% kadar kejayaan apabila padanan stereo penyesuaian serta dengan fungsi asas-Linear digunakan.

IMPROVED STEREO VISION ALGORITHMS FOR ROBOT NAVIGATION

ABSTRACT

The main motivation of this research is to find the best depth and direction for navigating a robot using stereo vision by solving the difficulties in finding disparity value for low information, noisy and tilted images as problem statement. An adaptive window method is implemented in three approaches. Approach 1 and 2 use common cost functions such as SSD, SAD and GB (gradient). Approach 3 uses a Linear-based function. By using adaptive method, the error in computing the disparity value for SSD is 12%, for Gradient-based is 10% and for Linear-based is 7%. SSD is 8 and 2 times faster than Gradient-based and Linear-based functions respectively. The linear-based technique with 50% more accurate than SSD is a suitable tool for stereo vision applications. The proposed denoising method for Lena and Barbara images are compared with ROF model by computing Total Variation (TV). When TV is 0.88 for noisy Lena image, ROF reduces TV to 0.21, Linear-based and Gradient-based decreases TV to 0.24 and 0.26 respectively. When TV is 0.88 for noisy Barbara image, TV is decreased to 0.68 by ROF, 0.62 by Linear-based and 0.65 by Gradient-based techniques. A stable platform system is designed and developed for stabilizing vision line horizontally for a moving robot. The system is equipped by a closed loop tilt and pan stabilizer using a dual-axis accelerometer sensor. It stabilizes the tilt and pan angles with 0.5 second time constant and steady state error lower that 0.025 radian. By using adaptive stereo matching that uses linear cost function, the quality of the best direction for navigation is improved to 88% of success rate.

CHAPTER 1
INTRODUCTION

1.1 Overview

The sense of vision has a considerable effect on human life. It allows us to have a proper understanding of the surrounding world. Man uses his vision to explore new unfamiliar environments and be aware of events occurring in remote distances. The optical vision system is an optical receiver sensor that maps the 3D information of the world to a 2D image on the retina plane. Due to the reduction of the dimension of the world information from 3 to 2, the depth of the objects in the scene is lost. Our natural Stereo vision mechanism, including our eyes, recovers the depths of the objects by two left and right images to present a better realization of the surrounding world. Imitating this natural system, the artificial stereo vision, consisting of minimum left and right images, estimates the depth of the objects in the scene via implementing the stereo algorithms including the correspondence and reconstruction processes. Image processing requires the introduction of assumptions on some of the unknown properties of the scene to extract the desired information. These assumptions depend on the properties of the applications that use the vision algorithms as a main tool.

Artificial stereo vision was found to be a very useful tool in many industrial and military applications after the improvement in the technology of digital cameras and fast computers. Vision systems have been used in many applications such as road detectors, traffic control, parking control, robot navigation, automatic landing and passive distance measurements **(**Ma et al., 2004).

1.2 Motivation

Stereo vision provides passive distance measurement. In contrast with active distance measuring systems like radars and lasers, stereo vision can estimate the depth of the target based on the left and right images passively. It can offer more information about the scene in relation to the shapes and motion of the objects. Stereo vision system can provide autonomous navigation mechanism for

robots and other moving machines like helicopters. The accuracy of the operation is directly related to the specification of the vision equipment and algorithms.

Improving the estimation of depth and direction for navigating a robot by stereo vision algorithms is the main motivation in this thesis.

1.3 Problem Statement

As in image formation that a 3D scene is projected onto a 2D image plane (Figure 1.1), the information about the depth is lost; image processing is an ill-posed problem. Another difficulty is when there is no information about the scene. The third, mostly in stereo vision, has become critical as there are some large smooth or noisy areas in the images. The last problem is very important where we want to navigate a moving vehicle by a stereo system. Computing disparity for areas without edge or smooth parts in stereo images is a big problem in stereo vision. The source of errors in stereo vision systems is as the follows: (Cyganek and Siebert, 2009)

- **i.** The errors including differences in cameras parameters (resolution, sampling rate and internal noise), inaccurate optical geometry configuration (mechanical vibrations and non calibrated cameras).
- **ii.** The errors including improper image processing algorithms (mostly denoising and deblurring) and stereo vision methods (matching, reconstruction and depth estimation).

The problems in stereo vision considered to be solved in this work are divided into four categories as follows:

- **i.** Difficulty in computing disparity value for areas without edge or smooth parts.
- **ii.** Difficulty in comparing two image windows that is the main action in stereo matching.
- **iii.** Error in disparity when images are noisy and blurred.
- **iv.** Error in disparity when stereo images are tilted by tilt and pan angles when a robot moves in rough terrain.

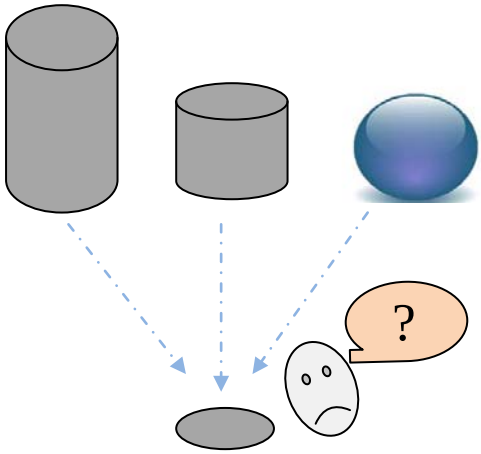


Figure1.1: Major problem in image formation

1.4 Research Objectives

The objectives in this thesis are as follows:

- **i.** To study and propose a denoising method applicable in stereo vision system for removing noise of stereo images, some approaches to solve difficulties in comparison of two image windows and find disparity for low information stereo images
- **ii.** To design and implement a stable platform system for the compensation of stereo images that are tilted by tilt and pan angles when robot moves in uneven terrain.
- **iii.** To implement the proposed methods in (**i**) together with the platform system in (**ii**) on an autonomous vehicle.

1.5 Scope of the Thesis

In navigating a robot by stereo vision, the most important problem is to find the best direction and depth to move the robot. To obtain this information, disparity map should be computed mainly for big and smooth parts of stereo images like a clean wall or door. In this case, because all the points are similar in intensity, disparity cannot be computed. By using and improving the latest techniques in stereo matching, an adaptive window stereo matching is proposed and implemented. The results are compared and evaluated to choose the best one.

To remove noise from stereo images, a denoising method applicable in robot navigation is proposed and implemented. The results are compared with energy functional noise removal algorithm ROF (Rudin et al., 1992).

When the robot moves in uneven terrain, stereo images are tilted by tilt and pan angles. This problem reduces accuracy of disparity and increases computation time. To remove this difficulty, a stable platform system will be designed and implemented to compensate the tilt and pan angle practically.

By using stable platform system and implementing proposed adaptive window stereo matching and denoising, the robot can find the best depth and direction from noisy images in general environment including big and smooth obstacles when it is moving in rough terrain.

Implementation time as an important subject in stereo vision is mainly contributed by the delay in capturing the stereo images by cameras and the limitations of matlab program and the operating system. The work does not cover on the improving the limitations of implementation time and delay in cameras and communication system. This work focuses on improving the accuracy of the navigation system especially when the robot moves in low information world and uneven terrains. The implementation time can be improved by using faster cameras and communication system and with using special hardware and software designed for image processing such as DSP image processor and C compiler.

1.6 Thesis Organization

The organization of this thesis is as follows:

Chapter 1 describes thesis overview, motivation, problem statement and objectives. In chapter 2, the related work on image denoising and stereo vision including traditional and improved methods will be discussed. Chapter 3 contains the methods proposed in this study including denoising, deblurring and edge detection in image processing and an adaptive window matching algorithm in stereo vision. Chapter 4 is about the stable platform system, along with the parts of the robot. Chapter 5 presents the results of the implementation of the methods in this work. Chapter 6 includes the thesis conclusions and future works.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

Literature review in this work is divided into three categories which are the related previous work in denoising algorithms, stereo matching methods and robot navigation. These categories are important subjects in robot navigation using stereo vision.

2.2 Denoising Algorithms

In image processing, denoising is one of the most important issues. It has a considerable role, not only in visual perception but also in other imaging techniques like stereo vision, deformable template-based segmentation and stereo reconstruction where the image quality has a direct influence on the performance of the evolution process. Total variation (TV) based denoising methods like ROF (Rudin et al., 1992) define an energy functional that preserves the edges of the image and smoothes the Gaussian noisy area. The TV regularization technique is a suitable method that can be extended to different noisy conditions. To have a high performance in denoising, the noise distribution model should be compatible with the noisy image. The first solution technique of TV functional using the gradient projection method was very slow due to the existence of nonlinear terms in the evolution process.

In the last few years, using kernel-based techniques in image denoising has improved the quality of noise removal results. The reason for the success of using the kernel function in image denoising is related to the nature of the image formation process (Seo et al., 2007). An image is a result of a reduction dimension process which transforms a 3D scene to a 2D image plane by projecting the intensities of the objects with the same direction onto a pixel. Considering the projection process as a linear process, two important factors are concluded:

i. For a central pixel in an image, it is assumed that the intensity value is a linear combination of the intensities of its neighboring pixels.

ii. The pixels near and similar in distance and intensity of the central pixel are more influential.

Based on the mentioned factors, the real (hidden) value of intensity for a pixel is estimated as a linear combination of the parameter of the neighboring pixels such that the coefficients are changed by location and intensity to modulate the hidden value by spatial and range parameters. Kernel functions are the most suitable tools for spatial and range weighting. The weighting operation can be done by a neighboring window in the image or by another image. The image used in kernel functioning is called the guidance image. One of the most popular approaches using the guidance image is the bilateral filter (He et al., 2010). In the bilateral filter, the output for a pixel is a weighted average of the nearby pixels where the weight is a function of the intensity or color in the guidance image. A general framework for the filter kernels using guidance images (He et al., 2010) is:

$$q_i = \sum_j Q_{ij} I(p_j) \tag{2.1}$$

where q_i is the output for the i^{th} pixel in the output image $\boldsymbol{q}$, Q_{ij} the kernel function , p_j the j^{th}pixel in the input image $\boldsymbol{I}$ and i and j the pixel indexes. In a simple case of bilateral filter the kernel Q_{ij} can be in the following form:

$$Q_{ij} = \frac{1}{A_i} exp\left(-\frac{(X_j - X_i)^2}{{\sigma_x}^2}\right) exp\left(-\frac{(I_j - I_i)^2}{{\sigma_I}^2}\right) \tag{2.2}$$

X is the pixel coordinate, A_i is the normalizing factor such that $\sum_j Q_{ij} = 1$, and the parameters σ_x and σ_I are the controller coefficients. In cases in which guidance image $\boldsymbol{I}$ and input image $\boldsymbol{p}$ are identical, the filter degrades to the original bilateral filter.

Other important kernel based methods are introduced as Data-adaptive Kernel Regression, Non-Local Means and Optimal Spatial Adaption (Seo et al., 2007). In the method of Kernel Regression, the data model is defined as:

$$t_i = z(x_i) + e_i \qquad \forall i = 1,2 \dots P \tag{2.3}$$

where t_i is the value of intensity of the i^{th} pixel as the central pixel in the i^{th} window wi, $\in_i$ a Gaussian zero mean noise, $z(.)$ the unknown distribution of the intensity in the image at the location x_i and P the number of pixels. The first term, in the Taylor series of $z(x_i)$ around location x_i , is the best approximation for the value of intensity of the i^{th} pixel. According to this assumption, the minimization of the following optimization in the neighboring window wi yields the value of interest for $z(x_i)$:

$$min \sum_{j \in wi} |t_i - z(x_i)|^2 \, K(x_i - x_j, y_i - y_j) \tag{2.4}$$

where the kernel function $K(x_i - xj, y_i - yj)$, mostly in Gaussian form, is used to weight the neighboring pixels based on their spatial and range values. The value for $z(x_i)$ will be:

$$\hat{z}(x_i) = \frac{\sum_{j \in wi} K\left(x_i - x_j, y_i - y_j\right) t_j}{\sum_{j \in wi} K(x_i - x_j, y_i - y_j)} \tag{2.5}$$

The Non-Local Means method is concluded from the fact that the value of each pixel is in relation with other parts of the images. This technique uses certain windows around the central pixel to estimate $z(.)$ according to the kernel based scheme. The method can be expressed as:

$$\hat{z}(xi) = \frac{\sum_{i \neq j} K\left(Y_{wi} - Y_{wj}\right) t_i}{\sum_{i \neq j} K\left(Y_{wi} - Y_{wj}\right)} \tag{2.6}$$

Y_{wi} and Y_{wj} are the stacked vectors related to the windows wi (central window) and wj . The function $K\left(Y_{wi} - Y_{wj}\right)$ is the kernel function in the following form:

$$K\left(Y_{wi} - Y_{wj}\right) = exp\left\{-\frac{\left(Y_{wi} - Y_{wj}\right)^T V^{-1} \left(Y_{wi} - Y_{wj}\right)}{{\sigma_w}^2}\right\} \tag{2.7}$$

Matrix V is a diagonal matrix as:

$$V = diag\{\dots K(x_{j-1} - x_j), K(0), K(x_{j+1} - x_j), \dots\} \quad (2.8)$$

and σ_w is a parameter that controls the degree of smoothing.

Optimal Spatial Adaptive is an adaptive scheme of the Non-Local Means method which adaptively controls the degree of filtering. The basic strategy is iteratively changing the size of the neighboring windows to the size that satisfies an optimum condition (minimum energy difference between original and evolved image). The form of the kernel function is similar to Non-Local Means when matrix V contains the harmonic means of estimated local noise variance.

$$V = diag\left\{\dots, \frac{{\sigma_{i-1}}^2 {\sigma_{j-1}}^2}{{\sigma_{i-1}}^2 + {\sigma_{j-1}}^2}, \frac{{\sigma_i}^2 {\sigma_j}^2}{{\sigma_i}^2 + {\sigma_j}^2}, \dots\right\} \quad (2.9)$$

where ${\sigma_i}^2$ is the variance of noise in the window wi.

Another novel state of the art method was recently introduced as guided filter (He et al., 2010). The filter uses spatial and range weighting in a linear strategy. In a guided filter, output image $\boldsymbol{q}$ is the result of a linear operation on input image $\boldsymbol{p}$ such that the coefficients are computed by guidance image $\boldsymbol{I}$ (Figure 2.1). All pixels in the k^{th} neighboring window w_k in output image $\boldsymbol{q}$ are in a linear relation with their correspondencing pixels in guidance image $\boldsymbol{I}$ with the constant parameter a_k and b_k so that:

$$q_i = a_k I_i + b_k \qquad \forall\, i \in w_k \quad (2.10)$$

Coefficients a_k and b_k are computed for the current window w_k in a process of minimizing the least square error between output image $\boldsymbol{q}$ and input image $\boldsymbol{p}$ for the local window such that:

$$J_k = \sum_{i \in w_k} ((a_k I_i + b_k - p_i)^2 + \in {a^2}_k) \quad (2.11)$$

The parameter $0 < \in < 1$ is used to reduce the value of a_k . After minimization of the cost function J_k we have:

$$a_k = \frac{\frac{1}{|W|}\sum_i I_i p_i - \bar{I}\bar{p}}{(\sigma^2 + \in)} \quad (2.12)$$

$$b_k = \bar{p} - a_k \bar{I} \quad (2.13)$$

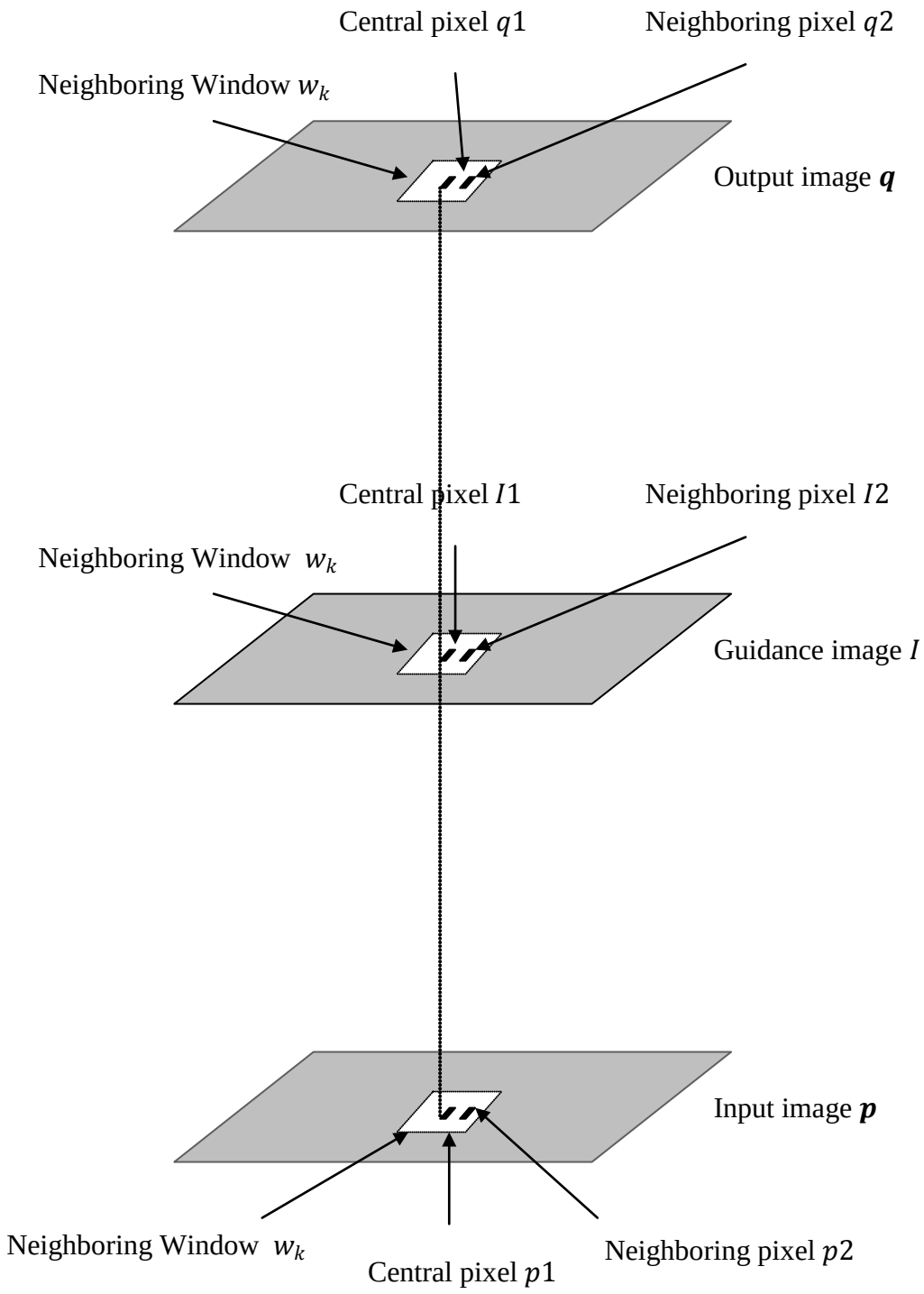


Figure 2.1: Images in a guided filter

where $|W|$ is the size of the window, $\bar{I}$ and $\bar{p}$ the mean values of guidance image $\boldsymbol{I}$ and input image $\boldsymbol{p}$ respectively and σ^2 is the variance of the image $\boldsymbol{I}$. All the parameters are computed in current window w_k . By moving the window through the images for each pixel in output image $\boldsymbol{q}$, $|W|$ different numbers a_k and b_k will be computed. So averaging the parameters yields the final coefficients for each pixel:

$$q_i = \bar{a}\ I_i + \bar{b} \tag{2.14}$$

where $\bar{a}$ and $\bar{b}$ are:

$$\bar{a} = \frac{1}{|W|}\sum_k a_k \tag{2.15}$$

$$\bar{b} = \frac{1}{|W|}\sum_k b_k \tag{2.16}$$

2.3 Stereo Matching Methods

This section reviews the stereo matching methods. Two comprehensive reviews on the subject are the papers that classify the methods and compare them (Brown et al., 2003), (Scharstein and Szeliski, 2002). The work in this thesis concentrates on stereo matching algorithms applicable for navigating moving vehicles. The heart of stereo matching is the process of image comparison which has been widely one of the most interesting subjects in computer vision (Scharstein and Szeliski, 2002).

The most basic step in stereo matching is computing the matching costs for a pair of pixels in stereo images. Using only a single intensity value has inappropriate disadvantages of which the most important causes are noisy images, differences in the parameters of the cameras, different distribution of intensity of light in the scene. In addition, the possibility of the equality between the intensities of two or more pixels in an image is highly expected. Aggregation scheme is a general solution in computer vision that uses neighboring pixels to compute the cost function. Neighboring pixels involved in the process are called the support region. In the stereo matching process, the most common cost functions are the

sum of squared differences (SSD) (Simoncelli et al., 1991) and the sum of absolute differences (SAD) (Kanade et al. 1995). Other traditional matching functions include binary matching cost functions (Marr and Poggio, 1976) based on binary features (Canny, 1986), sign of Laplacian function (Nishihara, 1987), gradient based measures (Scharstein, 1994) rank and census transforms as non-parametric criteria (Zabih and Woodfill, 1994) and phase and filter-bank responses (Jones and Malik, 1992).

In the aggregation step, each pixel is assigned to its feature that is computed in the related support region. A support region can be either two or three dimensional. A two-dimensional case uses a square window around the pixels in an x-y space while a three-dimensional support region considers the value of current disparity as the third coordinate in an x-y-d space (Bobik and Intille, 1999). A simple support region is a set of neighboring windows in different sizes which can be convolved by a suitable kernel function (Hirschmuller et al., 2002). A low pass filtering is one of the simplest cases.

Aggregation is the basic step in two main stereo matching methods: Local and Global. In local methods, disparity is obtained in the minimum point of a cost function which is computed for the difference between a reference window in one image and a test window of the same size in another image. Local methods are divided into feature and area matching. In contrast to local matching methods using small-sized windows, global methods find disparity by using all cost functions in an optimum process. The most important global methods include Belief Propagation (Tappen and Frreeman, 2003), Dynamic Programming (Cormen et al., 2001), Nonlinear Diffusion (Scharstein and Szeliski, 2002), Tensor Voting (Mordohai and Medioni, 2006) and Graph Cuts (Kolomogrove and Zabih, 2005). Figure 2.2 shows a diagram of local and global methods.

The local methods are (Cyganek and Siebert, 2009):

i. **SAD:** Sum of Absolute Differences
ii. **SSD**: Sum of Squared Differences
iii. **NSSD**: Normalized Sum-Squared Differences
iv. **CV**: Covariance-Variance
v. **NSCP**: Normalized Sum of Cross Products
vi. **SD**: Statistical Distance

vii. **F-N**: Fuzzy logic and Neural network approaches

viii. **AW**: Adaptive Weighting

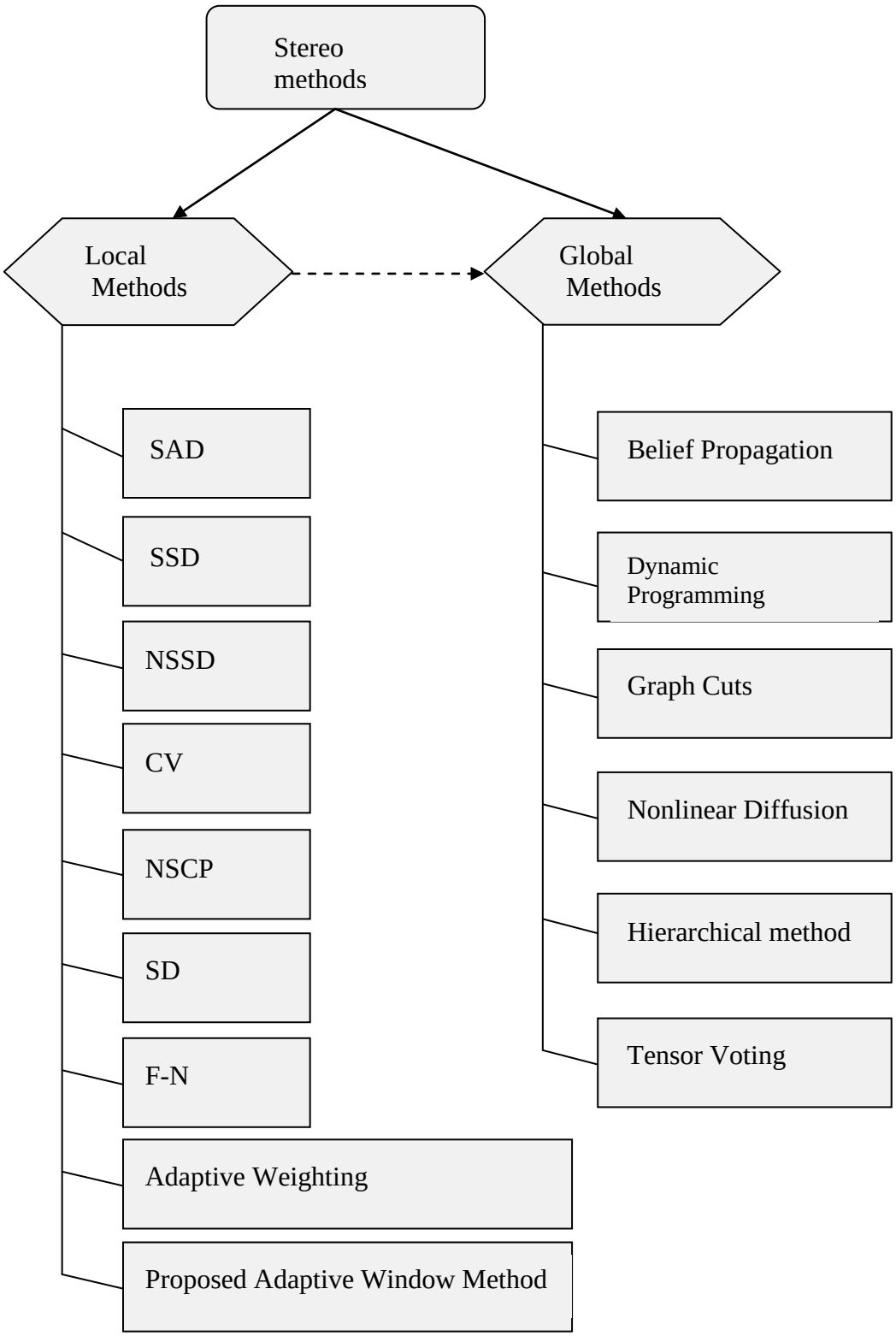


Figure 2.2: Stereo matching methods

The result of each stereo matching method is a disparity map. Based on the type of disparity maps, stereo methods can be divided into two groups: dense and sparse disparities (Figure 2.3). In dense disparity maps, all pixels are given a particular disparity value while in the sparse disparity maps, the disparity value is given to only some parts of the images. The dense map is more accurate while the sparse map is faster in computation (Scharstein, 1999).

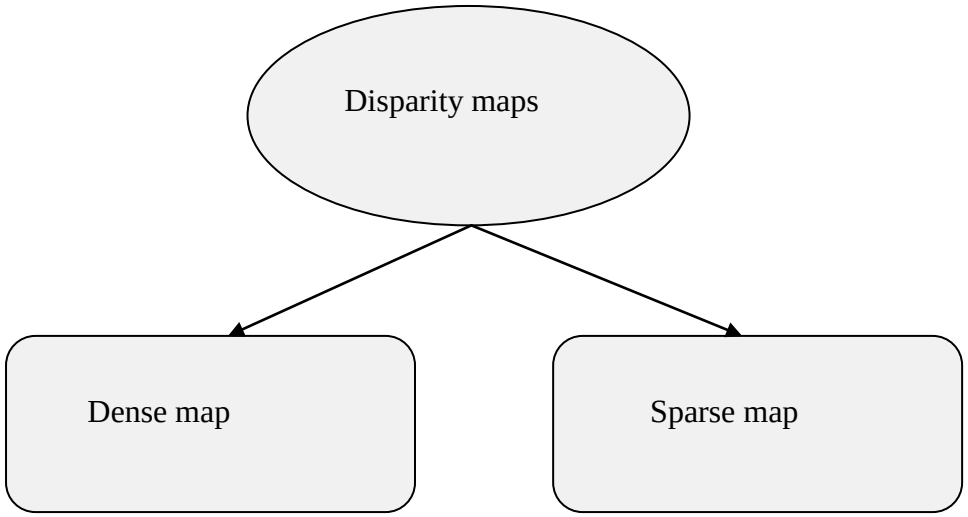


Figure 2.3: Diversity in disparity maps

Although stereo matching methods are different in many aspects, they can be carried out in four steps (Scharstein and Szeliski, 2002) shown in Figure 2.4. Disparity map post processing is to improve the quality of the resulting disparity. The main methods are the sub-pixel estimation of the disparity map, cross checking disparity verification, filtering the disparity map and interpolation of missing disparity values (Cyganek and Siebert, 2009).

Global methods represent a more accurate disparity map but the computation time is higher than local approaches. One of the disadvantages of local approaches is that in many cases, the real matching windows do not satisfy the minimum cost value scheme. Using some other characters in the windows to increase separation capability is the idea behind the introduced modification techniques.

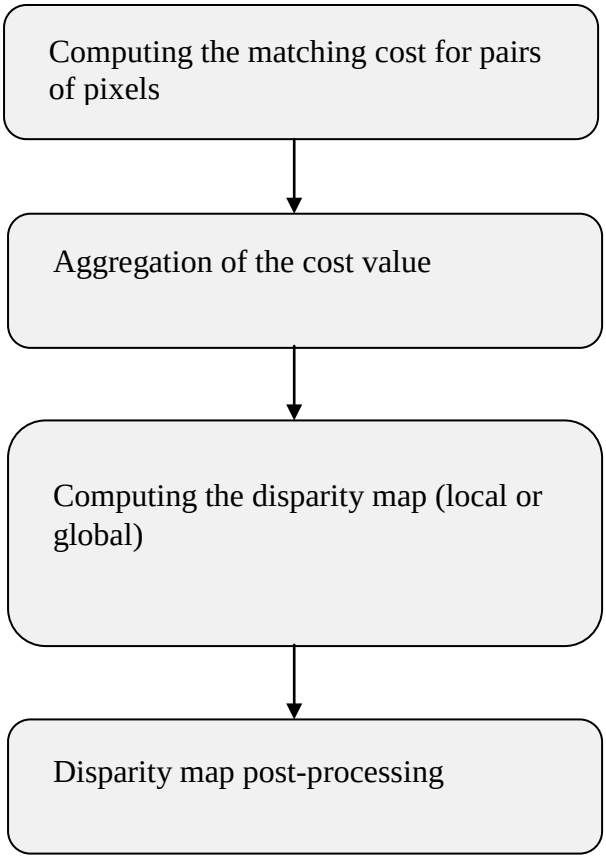


Figure 2.4: Basic steps for most stereo methods

Recently, some new local methods called adaptive-weight algorithms have been proposed. By preserving accuracy, they are fast and applicable in real time cases (De-Maeztu et al., 2011). It is shown that adding distance and statistical measures as weighting cost function for each pixel in the windows can modify local methods.

The first adaptive-weight algorithm (Yoon and Kweon, 2006) introduced a dissimilarity function in which the difference of intensities between two pixels in stereo windows was weighted by two other factors:

i. Euclidean distance between the values of intensities of two match pixels with central pixels on the windows.

ii. Spatial Euclidean distance between each pixel in the window and the central pixel.

This weighting increases the importance of the pixels near in location and similar in intensity with the central pixel. This adaptive weight algorithm is a bilateral filter. The output of this filter can be a general function of the characters of the central pixel and its neighboring pixels.

Other adaptive weight based methods join the histogram of the image with the cost function (Ju and Kang, 2009). To reduce the computation time in computing the histogram, a method called the integral histogram is proposed for faster computation of the histogram of difference window in a local stereo algorithm (Porikli, 2008). This method has a considerable role in increase of speed of the disparity map computation. According to the integral histogram technique, the histogram for a moving window uses the common part of the previously computed histogram. The cost function incorporates the values of the histogram of each pixel in the window as an additional weight (Ju and Kang, 2009). According to Figure 2.5 the first adaptive weight method (Yoon and Kweon, 2006) can be formulated according to the following relations:

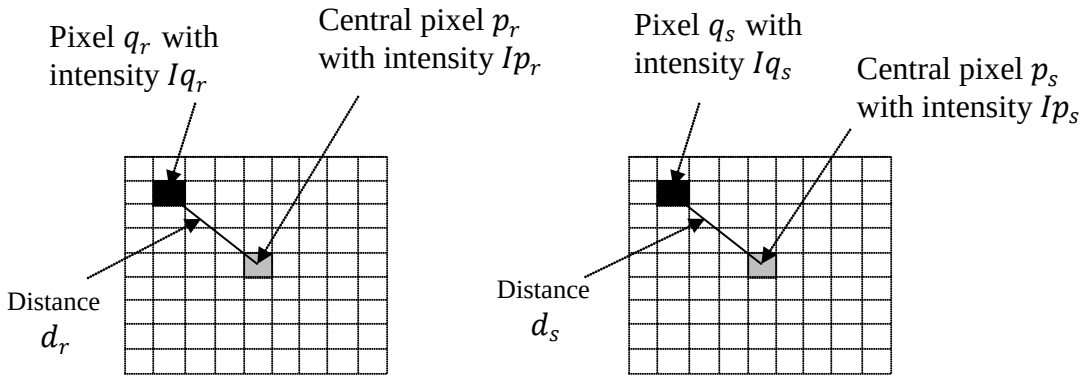


Figure 2.5: (a) Reference window (b) Search window

$$E(r,s) = \frac{\sum_{q_r \in wr, q_s \in ws} Q(Iq_r, Ip_r) Q(Iq_s, Ip_s) Q(d_r) Q(d_s) e(r,s)}{\sum_{q_r \in wr, q_s \in ws} Q(Iq_r, Ip_r) Q(Iq_s, Ip_s) Q(d_r) Q(d_s)} \tag{2.17}$$

The function $e(r,s)$ is a cost function like SSD. $E(r,s)$ is the dissimilarity function between the reference and search window. The function $Q(Iq_r, Ip_r)$ and

$Q(Iq_s, Ip_s)$ are the weight of the difference in the intensity value between a pixel and the central pixel in the reference and search window as:

$$Q(Iq_r, Ip_r) = e^{-\frac{(Iq_r - Ip_r)^2}{\sigma_1}} \tag{2.18}$$

$$Q(Iq_s, Ip_s) = e^{-\frac{(Iq_s - Ip_s)^2}{\sigma_1}} \tag{2.19}$$

The parameter σ_1 is for tuning and $Q(d_r)$ and $Q(d_s)$ are kernel functions:

$$Q(d_r) = e^{-\frac{(d_r)^2}{\sigma_2}} \tag{2.20}$$

$$Q(d_s) = e^{-\frac{(d_s)^2}{\sigma_2}} \tag{2.21}$$

Parameter σ_2 is for tuning and the function $e(r,s)$ is a cost function like SSD that projects the energy of the difference in intensity of the two corresponding pixels in stereo windows on to the dissimilarity function $E(r,s)$. The function $e(r,s)$ can be considered generally as:

$$e(r,s) = f(q_r, q_s) \tag{2.22}$$

where the function $f(q_r, q_s)$ is an arbitrary cost function. Implementation of function E is complex in time. The dissimilarity function (Ju and Kang, 2009) that uses integral histogram for faster computation is:

$$E(r,s) = \frac{\sum_{i \in N_H} Q(p_r, i) Hist(pr, i)}{\sum_{i \in N_H} Q(p_r, i)} \tag{2.23}$$

Function $Hist(pr, i)$ is the histogram function of the difference window (reference window minus search window), i the bin of histogram , N_H the number of bins and p_r the central pixel in the reference window. The weight $Q(p_r, i)$ is computed for the central pixel in the reference window and bin of the histogram. According

to the equation 2.7, the order of computation is independent of the size of the stereo windows and can be reduced by decreasing the number of the bins in the histogram.

Another method which uses the guided filter algorithm in local stereo matching is called linear stereo (De-Maeztu et al., 2011). The guidance image is a combination of the reference and target windows stacked together in vector $\begin{bmatrix} I_r \\ I_t \end{bmatrix}$. Output image O for current windows is defined as:

$$O = \boldsymbol{a} \begin{bmatrix} I_r \\ I_t \end{bmatrix} + b \tag{2.24}$$

The vector $\boldsymbol{a}$ is $1 \times n$ such that n is 2 or 6 for grayscale or color images respectively. The aggregation energy function to be minimized is:

$$J = \sum_{p \in w} \left(\left(\boldsymbol{a}_{1\times n} \begin{bmatrix} I_r \\ I_t \end{bmatrix}_{n\times 1} + b - e(p, q_d) \right)^2 + \in \boldsymbol{a}\boldsymbol{a}^T \right) \tag{2.25}$$

Where $e(p, q_d)$ is a cost function, computed for $I_r(p)$ and $I_t(q_d)$ as the reference and target windows respectively for the disparity d and central pixels p and q_d in the two windows. After minimization:

$$\boldsymbol{a} = (\Sigma_d + \in U)^{-1} \left(\frac{1}{|W|} \sum_p \begin{bmatrix} I_r \\ I_t \end{bmatrix} e(p, q_d) - \begin{bmatrix} \mu_r \\ \mu_t \end{bmatrix} \bar{e} \right. \tag{2.26}$$

$$b = \bar{e} - \boldsymbol{a} \begin{bmatrix} \mu_r \\ \mu_t \end{bmatrix} \tag{2.27}$$

Σ_d is the $n \times n$ covariance matrix related to vector $\begin{bmatrix} I_r \\ I_t \end{bmatrix}$, U a $n \times n$ identity matrix, $|W|$ the size of the windows, $\begin{bmatrix} \mu_r \\ \mu_t \end{bmatrix}$ a mean vector of $\begin{bmatrix} I_r \\ I_t \end{bmatrix}$ and $\bar{e}$ is the mean value of the intensities in the difference window. According to the previous illustration, the final value for coefficients is computed by averaging on the aggregation windows.

During the local matching process, if the two windows are similar (matching condition) the covariance matrix Σ_d is lower than U. Therefore the coefficient vector $\boldsymbol{a}$ tends toward zero and b tends toward $\bar{e}$. In contrast, when two reference and target windows are not matched, the parameters of vector $\boldsymbol{a}$ become close to 1. The parameter $\in$ controls the performance of this behavior.

Cost function $e(p, q_d)$ can be one of the traditional functions like SSD and SAD. Linear local stereo matching by guided filters is faster and more accurate than those using bilateral filters. The key point in guided filter is the averaging process to assign the coefficient parameters to each pixel. According to this action for a window with $s \times s$ number of pixels, the parameters are computed $2s - 1$ times for each pixel. This condition allows us to decrease the size of the searching window to $s = 3$x3 pixels (the minimum size with $r = 1$ neighboring radius) to have a faster response. This implementation is named $O(1)$ in the related literatures (De-Maeztu et al., 2011).

Navigating a vehicle by stereo vision method in a real time application should satisfy some factors. Implementation should be relatively fast and accurate. Although the disparity map is more accurate in global methods, the response of global methods is slow (De-Maeztu et al., 2011). As Local methods are fast in computing the disparity map with enough accuracy for the navigation of a vehicle, they are a better choice in stereo vision navigation system for autonomous vehicles. These new local methods can yield comparable responses with global algorithms with less computation time.

In many stereo vision applications including known patterns stereo algorithms try to find disparity for special objects. Some examples are stereo road detectors (Betrozzi and Broggi, 1998) and helicopter landing (Xu et al., 2006). In contrast with the special cases, in general conditions where an autonomous vehicle wants to move in an unknown environment, the difficulties in extracting all necessary information for motion still exist. Two main problems are noisy or smooth areas including areas with no information on disparity and finding the optimum depth and angle for navigation from the disparity map.

2.4 Mobile Robot Navigation

Navigation is a fundamental problem in mobile robotics. The local navigation problem deals with navigation on the scale of a few meters, where the main problem is obstacle avoidance. The global navigation problem deals with navigation on a larger scale in which the robot cannot observe the goal state from its initial position (Ulrich and J. Borenstein, 1998). For navigating a mobile robot in an environment, the surrounding information should be obtained to detect the obstacles. Identification of the surrounding environment requires spatial sensors. Three spatial sensors are Laser Range Finders, Sonar Sensors and Stereo Vision. Laser range finders are sensors that measure the distance by computing the time of flight of an emitted laser impulse. Sonar sensors detect obstacles based on transmitting and receiving sound waves but their resolution is lower than that of lasers. While Sonar and lasers are active systems, Stereo vision is a passive technique for inferring the three dimensional position of objects from two or more simultaneous images of a scene. Mobile robots can take advantages of a stereo vision system as a reliable and an effective way to extract range information from the surroundings. Generally, accuracy is adequate in depth estimation and obstacle detecting applications.

The most important advantage is that stereo vision is a passive sensing technique and there is no interference with other sensor devices in applications including multiple robots. In mobile robot navigation using stereo vision, stereo images require suitable pre and post-processing operation to estimate information of three dimensional ranges. Mobile robot navigation using obstacle detection is a method based on stereo depth measurement with no corresponding points (Kumano et al., 2000). The method is fast enough for mobile robot navigation. Distance transform methodology (DT) (Chin et al., 2001) is an approach for robot navigation. DT can be used in path planning for indoor robot navigation and also in performing obstacle avoidance simultaneously. There have been a lot of researches interested in a more intelligent design of autonomous controllers for controlling the basic flight modes of unmanned helicopters (Fang et al., 2008). Many researchers focus on the dynamic control problems (Wang and Yang., 2011). In the work of Katzourakis et al. (2009) and Xu et al. (2006), navigation and landing with the stereo vision system was discussed. Xu et al. used the stereo

vision system for estimating the position of the body. From the work of Xu, it was shown that the stereo vision does work for the position estimation.

2.5 Summary

The literature review presented in this chapter is about denoising methods, stereo matching algorithms and robot navigation.

In denoising, three basic frameworks are discussed which are energy functional based methods, kernel based algorithms and linear based techniques. In energy functional methods, ROF as the first model works based on finding the total variation (TV) norm minimizer. Kernel and linear based methods are two types of guided filtering algorithm. In kernel based methods Data-adaptive Kernel Regression, Non-Local Means and Optimal Spatial Adaption are most popular. They are mainly designed for denoising and super resolution applications. Linear based methods are similar to kernel based methods but they use spatial and range weighting in a linear strategy.

Stereo matching algorithms are divided into two local and global methods. In local stereo matching, the most common cost functions are the sum of squared differences (SSD) and the sum of absolute differences (SAD). The most important global methods include Belief Propagation, Dynamic Programming, Tensor Voting, and Graph Cuts. Recently some new local methods called Adaptive-Weight (AW) algorithms have been proposed that are fast and applicable in real time cases. AW uses the values of intensity and spatial Euclidean distance between each pixel in the window and the central pixel. Another method which uses the guided filter algorithm in local stereo matching is called linear stereo. Linear stereo is similar to AW in accuracy but faster in implementation.

Mobile robot navigation is divided into local and global types. In mobile robot navigation, stereo vision is a suitable technique for inferring the three dimensional position of objects from two or more simultaneous images of a scene. Occupancy grid mapping and potential field and transform methodology (DT) are two popular approaches for robot navigation.

CHAPTER 3
RESEARCH METHODOLOGIES

3.1 Introduction

The steps in classic methods in stereo robot navigation consist of local matching, cross checking disparity verification, filtering disparity map, interpolation of missing disparity values and computing the deepest part of the disparity map (Cyganek and Siebert, 2009). The complexity is increased for images with low information that consist of large, smooth, noisy or blurred parts. Classic stereo navigation, denoising and deblurring are complex in implementation. To make the process easy, fast and applicable in robot navigation, a denoising and a debluring method in image processing and an adaptive window algorithm in stereo vision are presented as solutions to these problems.

First in section 3.2.1 and section 3.2.2, two edge detector functions (gradient and linear) are introduced. Then, these functions are used in three proposed methods as denoising in section 3.2.3, edge detection in section 3.2.4 and deblurring in section 3.2.5. The methods are applicable in robot navigation. In section 3.3 an adaptive window method is proposed as a solution for low information stereo images. This method is implemented in three approaches. The two first approaches in section 3.3.2 and section 3.3.3 use common cost function such as SSD and SAD. In the third approach (section 3.3.4) a linear cost function is used to improve the results in stereo robot navigation. This linear cost function is illustrated in section 3.3.4. Section 3.4 is the summary of the chapter.

3.2 Methodologies in Image Processing

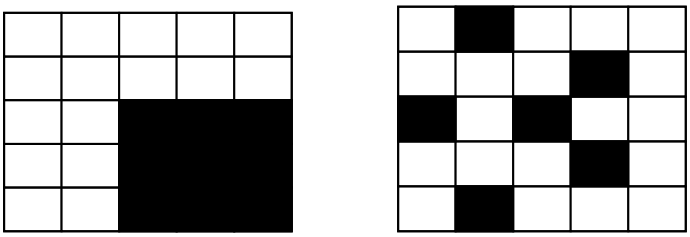
In this section two edge detector functions are introduced as gradient function and linear function. It is shown how an edge detector can be used as a separator function for image denoising. An edge detection and a deblurring technique using separator function are also suggested.

3.2.1 Gradient Based Edge Detector Function

A special gradient based edge detector function is considered by computing the absolute value of gradient of image u in a window ΩN with N number of pixels. By considering the windows in Figure 3.1, $h(u)$ is introduced as:

$$h(u) = \frac{|\sum_{\Omega N} \nabla_x u| + |\sum_{\Omega N} \nabla_y u|}{N} \quad (3.1)$$

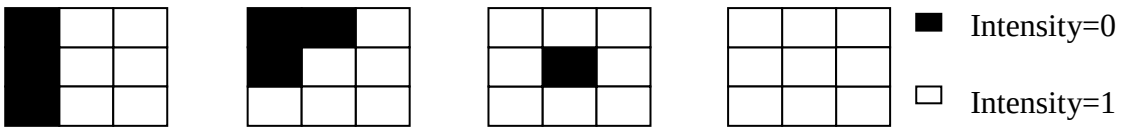
To explain how $h(u)$ works, a window ΩN with $N = 5\text{x}5$ pixels is considered in Figure 3.1. An example of a window containing an edge is shown in Figure 3.1(a). In this form of neighboring of pixels, $h(u)$ has a high value because there is a total difference between two levels of intensity in the window. For noisy or smooth configurations like Figure 3.1(b), $h(u)$ is low because the average of variation of intensity is small.



(a) (b)

Figure 3.1: (a) A typical edge window (b) Noisy window

A numerical example of $h(u)$ for four different 3x3 windows is shown in Figure 3.2.



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

Figure.3.2 A window including (a) Line (b) Edge (c) Noise (d) Smooth

Four values of $h(u)$ are $\frac{1}{3}, \frac{4}{9}, 0, 0$ respectively for the windows a, b, c and d. It can be seen that $h(u)$ is low for windows containing noisy or smooth areas.

3.3.2 Linear Based Edge Detector Function

The second function works based on the linear strategy used in guided filters. Considering a line of pixels inside window w (Figure 3.3) in image $\boldsymbol{q}$ for a central pixel located at the position (x_j, y_j), a linear relation can be written as:

$$a_1 U^+{}_j + a_2 U^-{}_j - mot_j = e_j \tag{3.2}$$

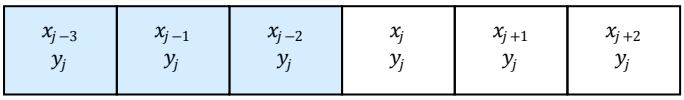


Figure 3.3: Line of pixels inside a window

where a_1 ,a_2 are the coefficients and e_j is the error in estimation and $U^+{}_j$, $U^-{}_j$ and mot_j are:

$$U^+{}_j = \sum_{i \gg j} u(x_{j+i}, y_j) \tag{3.3}$$

$$U^-{}_j = \sum_{i<j} u(x_{j+i}, y_j) \tag{3.4}$$

$$mot_j = U^+{}_j + U^-{}_j \tag{3.5}$$

After aggregating in the window:

$$a_1 \sum_{j \in w} U_j^+ + a_2 \sum_{j \in w} U_j^- - \sum_{j \in w} mot_j = \sum_{j \in w} e_j \tag{3.6}$$

The aggregated cost function is:

$$J = \sum_{j \in w} (a_1 U_j^+ + a_2 U^-{}_j - mot_j)^2 + \in (a_1{}^2 + a_1{}^2) \tag{3.7}$$

where the second term is added to prevent the increase of values of coefficients a_1 and a_2. $\in$ is a parameter that controls the dynamic response of the edge detector. This minimization scheme yields:

$$R = (V^T V)^{-1} V^T T \tag{3.8}$$

where the vectors R and T and the matrix V are:

$$R = [a_1 \quad a_2]^T \tag{3.9}$$

$$T_{2\times 1} = \begin{bmatrix} \sum_{j \in w} mot_j \sum_{j \in w} U_j^+ \\ \sum_{j \in w} mot_j \sum_{j \in w} U_j^- \end{bmatrix} \tag{3.10}$$

$$V = \begin{bmatrix} \in + \sum_{j \in w} {U_j^+}^2 & \sum_{j \in w} U_j^+ \sum_{j \in w} U_j^- \\ \sum_{j \in w} U_j^+ \sum_{j \in w} U_j^- & \in + \sum_{j \in w} {U_j^-}^2 \end{bmatrix} \tag{3.11}$$

For a noisy or smooth window in which $\sum_{j \in w} U_j^+ \approx \sum_{j \in w} U_j^-$, the coefficients a_1 and a_1 are approximately equal. If the window w contains an edge, then:

$$\sum_{j \in w} U_j^+ \neq \sum_{j \in w} U_j^- \tag{3.12}$$

In this condition, the coefficients are far apart. The following function is used as a measure of the intensity of the edge:

$$h(u) = |a_1 - a_2| \tag{3.13}$$

A typical behavior of function $h(u)$ is shown in Figure 3.4, with its intensity varying between zero and 1. The edge is distributed horizontally in an image window with size 3x3.