Depth Map and 3D Imaging Applications:

Algorithms and Technologies

Aamir Saeed Malik
Universiti Teknologi Petronas, Malaysia

Tae-Sun Choi
Gwangju Institute of Science and Technology, Korea

Humaira Nisar
Universiti Tunku Abdul Rahman, Perak, Malaysia

Information Science
REFERENCE

Managing Director: Lindsay Johnston
Book Production Manager: Sean Woznicki
Development Manager: Joel Gamon
Development Editor: Michael Killian
Acquisitions Editor: Erika Carter
Typesetters: Mackenzie Snader
Print Coordinator: Jamie Snavely
Cover Design: Nick Newcomer

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Copyright © 2012 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher. Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Depth map and 3D imaging applications: algorithms and technologies / Aamir Saeed Malik, Tae Sun Choi, and Humaira Nisar, editors.
p. cm.
Summary: "This book present various 3D algorithms developed in the recent years to investigate the application of 3D methods in various domains, including 3D imaging algorithms, 3D shape recovery, stereoscopic vision and autostereoscopic vision, 3D vision for robotic applications, and 3D imaging applications"-- Provided by publisher.
Includes bibliographical references and index.
ISBN 978-1-61350-326-3 (hardcover) -- ISBN 978-1-61350-327-0 (ebook) -- ISBN 978-1-61350-328-7 (print & perpetual access) 1. Algorithms. 2. Three-dimensional imaging. I. Malik, Aamir Saeed, 1969- II. Choi, Tae Sun, 1952- III. Nisar, Humaira, 1970- IV. Title: Depth map and three-D imaging applications.
QA9.58.D47 2012
621.36'7015181--dc23

2011031955

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Chapter 21
Stereo Vision Depth Estimation Methods for Robotic Applications

Lazaros Nalpantidis
Royal Institute of Technology (KTH), Sweden

Antonios Gasteratos
Democritus University of Thrace, Greece

ABSTRACT

Vision is undoubtedly the most important sense for humans. Apart from many other low and higher level perception tasks, stereo vision has been proven to provide remarkable results when it comes to depth estimation. As a result, stereo vision is a rather popular and prosperous subject among the computer and machine vision research community. Moreover, the evolution of robotics and the demand for vision-based autonomous behaviors has posed new challenges that need to be tackled. Autonomous operation of robots in real working environments, given limited resources requires effective stereo vision algorithms. This chapter presents suitable depth estimation methods based on stereo vision and discusses potential robotic applications.

INTRODUCTION

Stereo vision is a reliable tool in order to exploit depth data from a scene, apart the pictorial one. The accuracy of the results depends on the choice of the stereo camera system and the stereo correspondence algorithm. Stereo correspondence is a flourishing field, attracting the attention of many researchers (Forsyth & Ponce, 2002; Hartley & Zisserman, 2004). A stereo correspondence algorithm matches pixels of one image (reference) to pixels of the other image (target) and returns

DOI: 10.4018/978-1-61350-326-3.ch021

Copyright © 2012, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

the corresponding vertical displacement as the reference pixel's disparity, which is proportional to its depth. Thus, stereo vision is able to retrieve the third dimension of a scenery and, therefore, its importance is obvious in issues such as traversability estimation, robot navigation, simultaneous localization and mapping (SLAM), as well as in many other aspects of production, security, defense, exploration and entertainment.

Stereo correspondence algorithms can be grouped into those producing sparse and those giving dense output. Feature based methods stem from human vision studies and are based on matching segments or edges between two images, thus resulting in a sparse output. This disadvantage is counterbalanced by the accuracy and the speed of calculations. However, robotic applications demand more and more dense output. This is the reason why most of the relevant literature is focused on stereo correspondence algorithms that produce dense output. In order to categorize and evaluate them a context has been proposed (Scharstein & Szeliski, 2002). According to this, dense matching algorithms are classified in local and global ones. Local methods (area-based) trade accuracy for speed. They are also referred to as window-based methods because disparity computation at a given point depends only on intensity values within a finite support window. Global methods (energy-based) on the other hand are more time consuming but very accurate. Their goal is to minimize a global cost function, which combines data and smoothness terms, taking into account the whole image. Of course, there are many other methods that are not strictly included in either of these two broad classes. A detailed taxonomy and presentation of dense stereo correspondence algorithms can be found in (Scharstein & Szeliski, 2002). Additionally, the recent advances in the field as well as the aspect of hardware implementable stereo algorithms are covered in (Nalpantidis, Sirakoulis, & Gasteratos, 2008b).

ISSUES OF ROBOTICS-ORIENTED STEREO VISION

While a heavily investigated problem, stereo correspondence is far from being solved. Furthermore, the recent advances in robotics and related technologies have placed more challenges and stricter requirements to the issue. However, common problems related to outdoor exploration, such as possible decalibration of the stereo system and tolerance to non-perfect lighting conditions, have been barely addressed. Robotic applications demand stereo correspondence algorithms to be able to cope with not ideally captured images of the working environments of the robots (see Figure 1) and at the same time to be able to provide accurate results operating in real-time frame rates. Some of the open issues of robotics-oriented stereo vision methods are the handling of non-ideal lighting conditions, the requirement for simple calculation schemes, the use of multi-view stereo systems, the handling of miscalibrated image sensors, and the introduction of new biologically inspired methods to robotic vision.

Non-Ideal Lighting Conditions

The correctness of stereo correspondence algorithms' depth estimations is based on the assumption that the same feature in the two stereo images should have ideally the same intensity. However, this assumption is often not valid. Even in the case that the gains of the two cameras are perfectly tuned, so as to result in the same intensity for the same features in both images, the fact that the two cameras shoot from a different pose, might result in different intensities for the same point, due to shading reasons. In general, stereo image pairs captured in real life environments often suffer from differentiations in illumination, as those shown in Figure 2. Moreover, in real environments, which is the case for robotic applications, the illumination is far from being ideal (Klancar, Kristan, & Karba, 2004; Hogue, German, & Jenkin, 2007).

Figure 1. Robots equipped with stereo cameras in a real environments

The issue is usually treated by using robust pixel dissimilarity measures, which are able to compensate for lightness differentiations. Such a measure is the Zero mean Normalized Cross-Correlation (ZNCC) (Binaghi, Gallo, Marino, & Raspanti, 2004; Corke, 2005). However, this measure's computation is rather demanding and indirect approaches have been employed (Sun & Peleg, 2003). On the other hand, Ogale and Aloimonos in (Ogale & Aloimonos, 2005a, 2005b, 2007) propose and use a compositional approach to unify many early visual modules such as segmentation, shape and depth estimation, occlusion detection and local signal processing. As a result this method can process images with contrast, among others, mismatches. The first-stage dissimilarity measure used in this method is the phase differences from various frequency channels. Apart from these dissimilarity measures, a luminosity-compensated dissimilarity measure (LCDM) has been proposed in (Nalpantidis & Gasteratos, 2010b) and will be discussed in detail in a following section of this chapter.

Simplicity of Computations

Autonomous robots rely on their own decision-making algorithms (De Cubber, Doroftei,

Figure 2. A stereo image pair suffering from illumination differentiations

Nalpantidis, Sirakoulis, & Gasteratos, 2009). In the case of stereo vision-based navigation, the accuracy and the refresh rate of the computed disparity maps are the cornerstone of its success (Schreer, 1998). The most urgent constraint in autonomous robotics is the real-time operation and, consequently, such applications usually utilize local algorithms (Labayrade, Aubert, & Tarel, 2002; Soquet, Aubert, & Hautiere, 2007; Kelly & Stentz, 1998; Zhao, Katupitiya, & Ward, 2007; Konolige et al., 2006; Agrawal, Konolige, & Bolles, 2007). The hardware implementation of already proposed algorithms found in literature is not always straightforward (Nalpantidis et al., 2008b). Nevertheless, the hardware implementation of efficient and robust stereo algorithms able to provide real-time frame rates, especially in the case of moving robots, is very appealing. The allure of hardware implementations is that they easily outperform the algorithms executed on a general-purpose computer and, thus, the achieved frame-rates are generally higher. Furthermore, the power consumed by a dedicated hardware platform, e.g. ASIC or FPGA, is considerably lower than that of a common microprocessor and the computational power of the robot's onboard available computers is left intact.

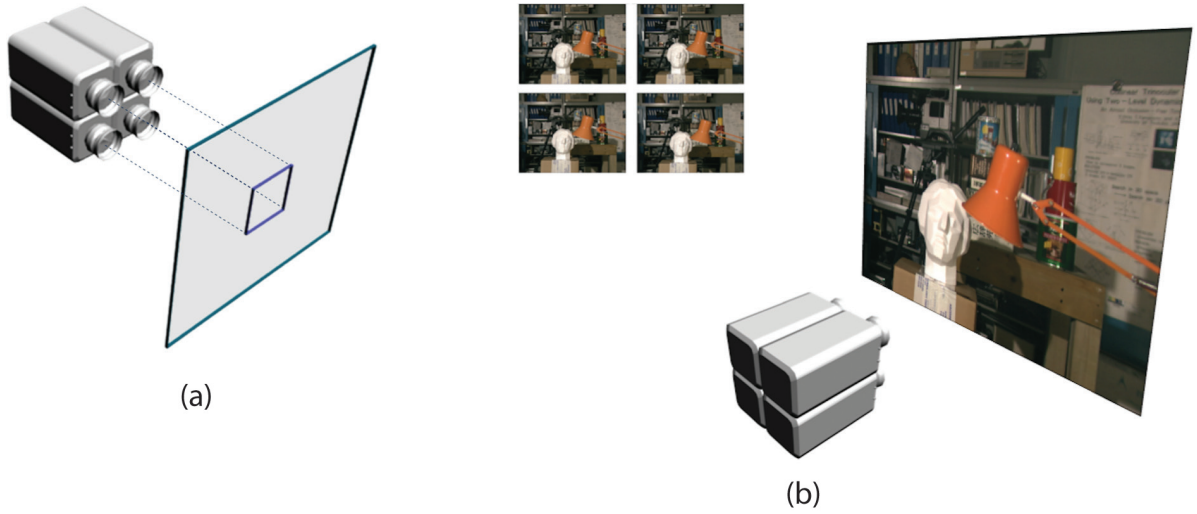
Multi-View Stereo Vision

Early previous work focused on developing stereo algorithms mostly for binocular camera configurations. However, redundancy can lead to more accurate and reliable depth estimations. More recently, due to significant boost of the available computational power, vision systems using multiple cameras are becoming increasingly feasible and practical. The transition from binocular to multi-ocular systems has the advantage of potentially increasing the stability and accuracy of depth calculations. The continuous price-reduction of vision sensors allowed the development of multiple camera arrays ready for use in many applications. For instance, Yang et al. (Ruigang, Welch, & Bishop, 2002) used a five-camera system for real-time rendering using modern graphics hardware, while Schirmacher et al. (Schirmacher, Li, & Seidel, 2001) increased the number of cameras and built up a six-camera system for on-the-fly processing of generalized Lumigraphs. Moreover, developers of camera arrays have expanded their systems so as to use tens of cameras, such as the MIT distributed light field camera (Yang, Everett, Buehler, & Mcmillan, 2002) and the Stanford multi-camera array (Wilburn, Smulski, Lee, & Horowitz, 2002). These systems are using 64, and 128 cameras respectively. Most of the aforementioned camera arrays are utilized for real-time image rendering. On the other hand, a research area that could also be benefited by the use of multiple camera arrays is the so-called cooperative stereo vision; i.e., multiple stereo pairs being considered to improve the overall depth estimation results. To this end, Zitnick (Zitnick & Kanade, 2000) presented an algorithm for binocular occlusion detection and Mingxiang (Mingxiang & Yunde, 2006) expanded it to trinocular stereo.

The system proposed in (Nalpantidis, Chrysostomou, & Gasteratos, 2009) is a combination of quad-camera sensor hardware and a custom-tailored software algorithm. The sensory configuration of the presented system consists of four identical cameras. The four cameras are placed so as their optical axes to have parallel orientation and their principal points to be co-planar, residing on the corners of the same square, as shown in Figure 3(a). The images captured by the upper-left camera are considered as the reference images of each tetrad. Each one of the other three cameras produces images to be corresponded to the reference images. Thus, for each tetrad of images three, differently oriented, stereo pairs result, i.e. an horizontal, a vertical and a diagonal one. The concept, as well as the result of such a group of cameras is presented in Figure 3(b).

The hardware configuration, i.e. the four cameras' formation, produces three stereo image pairs.

Figure 3. (a) The quad-camera configuration and (b) the results (up-left) and scene capturing (right) using a quad-camera configuration



Each pair is submitted to a simple and rapid stereo correspondence algorithm, resulting, thus, in a disparity map. For each disparity map a certainty map is calculated, indicating each pixel's reliability. Finally, the three disparity maps are fused, according to their certainties for each pixel. The outcome is a single disparity map that incorporates the best parts of its producing disparity maps. The percentage of pixels whose absolute disparity error is greater than 1 in the non-occluded, all, and near discontinuities and occluded regions using the aforementioned method are 10.8%, 12.6%, and 31.5% respectively. These results are significantly better than the results that would have been obtained if only a pair of images were to be used. The combined hardware and software system is able to produce accurate dense depth maps in frame rate suitable for autonomous robotic applications.

Uncalibrated Stereo Images

The issue of processing uncalibrated images is common to applications where the sensory system is not explicitly specified. The plethora of computations most commonly require the massive parallelization found in custom tailored hardware implementations. Moreover, the contemporary powerful graphics machines are able to achieve enhanced results in terms of processing time and data volume.

A hierarchical disparity estimation algorithm implemented on programmable 3D graphics processing unit is reported in (Zach, Karner, & Bischof, 2004). This method can process either rectified or non-rectified image pairs. Bidirectional matching is utilized in conjunction with a locally aggregated sum of absolute intensity differences (SAD). This implementation, on an ATI Radeon 9700 Pro, can achieve up to 50 fps for 256 × 256 pixel input images. The FPGA implementation presented in (Jeong & Park, 2004) uses the dynamic programming search method on a Trellis solution space. It copes with the vergent cameras case, i.e. cameras with optical axes that intersect arbitrarily, producing non-rectified stereo pairs. The image pairs received from the cameras are initially rectified using linear interpolation and then, during a second step, the disparity is calculated. The architecture has the form of a linear systolic array using simple processing elements. The design is canonical and simple to

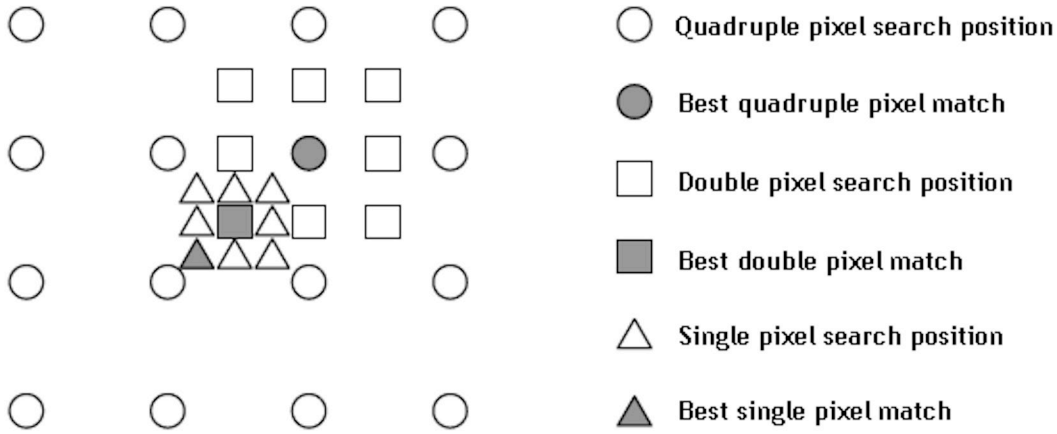
be implemented in parallel. The resulting system can process 1280×1000 pixel images with up to 208 disparity levels at 15 fps. An extension of the previous method is presented in (Park & Jeong, 2007). The main difference is that information from previously processed lines are incorporated so as to enforce better inter-scanline consistency. The running speed is 30 fps for 320×240 pixel images with 128 disparity levels. The number of utilized processing elements is 128. The percentage of pixels with disparity error larger than 1 in the non-occluded areas is 2.63, 0.91, 3.44, and 1.88 for the Tsukuba, Map, Venus and Sawtooth image sets, respectively. Finally, (Masrani & MacLean, 2006) proposes the utilization of a local weighted phase-correlation method. The platform used is the Transmogrifer-4 system containing four Altera Stratix S80 FPGAs. The system performs rectification and left-right consistency check to improve the accuracy of the results. The speed for 640×480 pixel images with 128 disparity levels is 30 fps.

The stereo vision algorithm presented in (Nalpantidis, Amanatiadis, Sirakoulis, & Gasteratos, in press) is inspired by motion estimation techniques. It is based on a fast-executed SAD core for correspondence search in both directions of the input images. The results of this core are enhanced using sophisticated computational techniques; Gaussian weighted aggregation and 3D cellular automata (CA) rules are used. The hierarchical iteration of the basic stereo algorithm is achieved using a fuzzy scaling technique. The aforementioned characteristics provide results of improved quality, being at the same time easy to be hardware implemented. Consequently, the presented algorithm is able to cope with uncalibrated input images. The presented scheme is block search-based and does not perform scanline pixel matching. As a result, it does require neither camera calibration nor image rectification. However, it is clear that block search approaches require more computational resources since the number of pixels to be considered is greatly increased. In order to address this problem, the presented algorithm employs a variation of a motion estimation algorithm (Yin, Tourapis, Tourapis, & Boyce, 2003), which is used for JVT/H.264 video coding (Wiegand, Sullivan, Bjntegaard, & Luthra, 2003). The adaptation of compression motion estimation algorithms into disparity estimation schemes can be effective both in accuracy and complexity terms, since compression algorithms also attempt to achieve complexity reduction while maintaining coding efficiency. On the other hand, CA have been employed as a intelligent and efficient way to refine and enhance the stereo algorithm's intermediate results. Let the maximum expected horizontal disparity for a stereo image pair be D. The dimensions of the stereo pixel matching search block are $D \times D$. For each search block, the disparity value is determined by the horizontal distance of the (single pixel sampling) best match in terms of minimum SAD, as shown in Figure 4.

In the first stage, the disparity algorithm finds the best match on the quadruple sample grid (circles). Then, the algorithm searches the double pixel positions next to this best match (squares) to assess whether the match can be improved and if so, the single pixel positions next to the best double pixel position (triangles) are then explored. The general scheme of the presented hierarchical matching disparity algorithm between a stereo image pair is shown in Figure 5.

Each of the intermediate disparity maps of the first two steps is used as initial conditions for the succeeding, refining correspondence searches. In order to perform the hierarchical disparity search three different versions of the input images are employed and the stereo correspondence algorithm is applied to each of these three pairs. The quadruple search step is performed as a normal pixel-by-pixel search, on a quarter-size version of the input images. That is, each of the initial images has been down-sampled to 25% of their initial dimensions. The quadruple search is performed by applying the stereo correspondence algorithm in $(D/4) \times (D/4)$ search regions, on the down-sized image pair (D being the maximum

Figure 4. Quadruple, double and single pixel sample matching algorithm

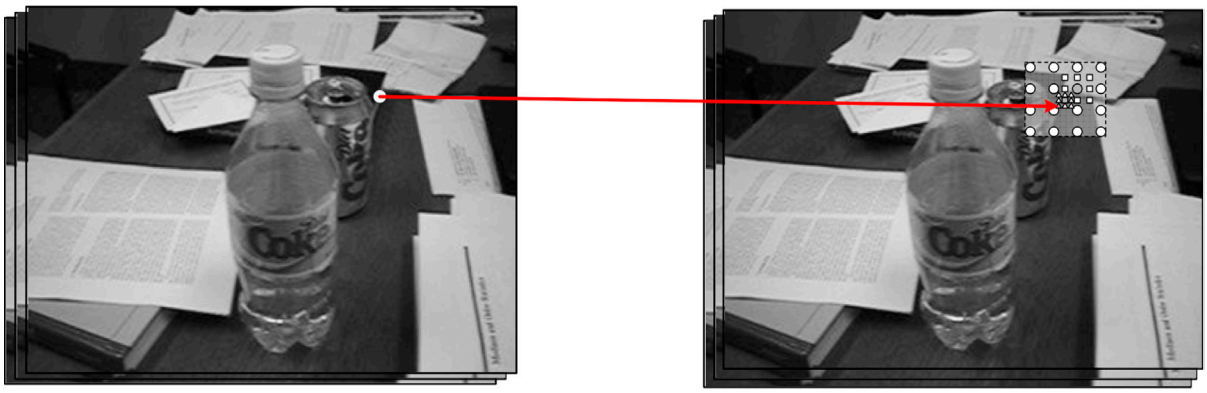


expected horizontal disparity value in the original image pair). The choice of the maximum searched disparity *D/4* is reasonable as the search is performed on a 1/4 version of the original images. This method provides good depth estimation with limited calculations, even for not calibrated input. The experimentally calculated Normalized Mean Square Error (NMSE) for radially distorted (10%) images of the Tsukuba, Venus, Teddy, and Cones data sets is 0.0712, 0.0491, 0.1098, and 0.0500 respectively. These results were in all cases less than 0.112% different from the results obtained for the original non-distorted image pairs and show that the discussed algorithm is robust against input's miscalibrations and distortions.

Biologically Inspired Methods

The success of the human visual system (HVS) in obtaining depth information from two 2D images still remains a goal to be accomplished by machine vision. Incorporating procedures and features from HVS in artificial stereo-equipped systems, could improve their performance. The key concept behind this transfer of know-how from nature to engineering is identifying, understand-

Figure 5. General scheme of the presented hierarchical matching disparity algorithm. The search block is enlarged for viewing purposes



ing and expressing the basic principles of natural stereoscopic vision, aiming to improve the state-of-the-art in machine vision. These principles are mainly involved in the aggregation step that most existing algorithms employ.

HVS has been studied by many branches of the scientific community. Physics have expressed color information through color spaces, while biology has investigated the response of the eyes to it and the physiology of the eye. Psychophysics has studied the relationship between individual stimuli's changes and the perceived intensity, which is applicable to vision as well as all the other modalities. On the other hand, the gestalt school of psychology suggested grouping as the key for interpreting human vision.

Gestalt is a movement of psychology that deals with perceptual organization. Gestalt psychology examines the relationships that bond individual elements so as to form a group (Forsyth & Ponce, 2002). As a consequence, a pattern emerges instead of separate parts. This pattern has generally completely different characteristics to its parts. Some of the gestalt rules by which elements tend to be associated together and interpreted as a group are the following:

- **Proximity:** elements that are close to each other.
- **Similarity:** elements similar in an attribute.
- **Continuity:** elements that could belong to a smooth larger feature.
- **Common fate:** elements that exhibit similar behavior.
- **Closure:** elements that could provide closed curves.
- **Parallelism:** elements that seem to be parallel.
- **Symmetry:** elements that exhibit a larger symmetry.

Gestalt laws have proven themselves to be precious tools in interpreting the way the human perceives his environment through vision. While all the laws are valuable in order to understand the context of an image, basic image processing tasks could be restricted to using the most basic ones. In order to express an image processing task through the prism of the gestalt theory, pixels should be considered as the elements. The correlation degree between them should be treated as the bonding relationship of the elements. The basic but at the same time important gestalt laws of proximity, similarity and continuity can then be applied in order to perform the given task.

As far as machine stereo vision is concerned, biological and psychological findings can be incorporated in the expression of proper correlation functions. Real life is the ultimate resource for finding right solutions in many fields of robotics, computer science and electronics (Mead, 1990; Shimonomura, Kushima, & Yagi, 2008; Berthouze & Metta, 2005). The natural selection process is a strict judge that favors the more effective solutions for each problem. Applying ideas borrowed from other sciences in technological problems can lead to very effective results. Consequently, further blending of biological and psychological findings with computer vision indicates a promising direction towards simple and accurate computer vision algorithms.

DEPTH MAPS COMPUTATION

The majority of stereo correspondence algorithms can be described using more or less the same structural set (Scharstein & Szeliski, 2002; Nalpantidis et al., 2008b). The basic building blocks are:

1. Computation of a matching cost function for every pixel in both the input images.
2. Aggregation of the computed matching cost inside a support region for every pixel in each image.
3. Finding the optimum disparity value for every pixel of one picture.
4. Refinement of the resulted disparity map.

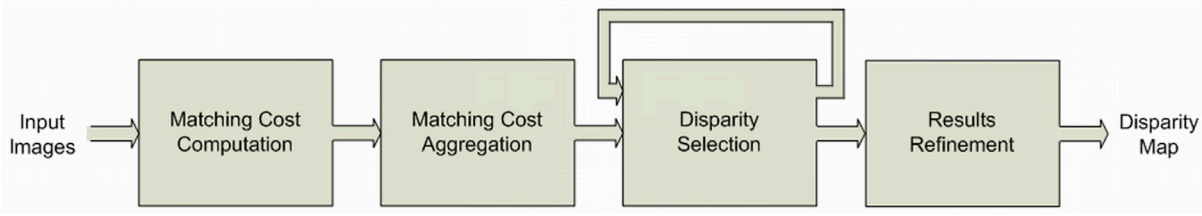
Every stereo correspondence algorithm makes use of a matching cost function in order to establish correspondence between two pixels. The results of the matching cost computation comprise the disparity space image (DSI). DSI is a 3D matrix containing the computed matching costs for every pixel and for all its potential disparity values (Muhlmann, Maier, Hesser, & Manner, 2002). Usually, the matching costs are aggregated over support regions. These regions could be 2D or even 3D (Zitnick & Kanade, 2000; Brockers, Hund, & Mertsching, 2005) ones within the DSI cube. Due to the aggregation step it is not single pixels that will be matched, but image regions. Aggregation of the matching cost values is a common and essential technique in order to suppress the effect of noise that usually leads to false matching. The selection of the optimum disparity value for each pixel is performed afterwards. It can be a simple winner-takes-all (WTA) process or a more sophisticated one. In many cases it is an iterative process as depicted in Figure 6. An additional disparity refinement step is frequently adopted. It is usually intended to interpolate the calculated disparity values, giving sub-pixel accuracy or assign values to not calculated pixels. The general structure of the majority of stereo correspondence algorithms is shown in Figure 6.

Given that iterative methodologies are generally not suitable for robotic applications, due to computation time restrictions, the main differentiations among the robotics-oriented covered algorithms have to do with the dissimilarity measure and the dissimilarity measure's aggregation scheme that they employ.

Dissimilarity Measures

Detecting conjugate pairs in stereo images is a challenging research problem known as the correspondence problem, i.e., to find for each point in the left image, the corresponding point in the right one (Barnard & Thompson, 1980). To determine these two points from a conjugate pair, it is necessary to measure the (dis-)similarity of the points. The point to be matched without any ambiguity should be distinctly different from its surrounding pixels. Several algorithms have been proposed in order to address this problem. However, every algorithm makes use of a matching cost function so as to establish correspondence between two pixels. The matching cost function is a measure that quantitatively expresses how much dissimilar (or equivalently similar) two image pixels are. There is a number of such measures that have been used in robotic vision algorithms, e.g. the absolute intensity differences (AD), the squared intensity differences (SD), the zero normalized cross correlation (ZNCC), phase-based measures, and the luminosity-compensated dissimilarity measure (LCDM). Each one of them has its merits and disadvantages regarding computational complexity and lighting-differentiations tolerance. An evaluation of various matching costs can be found in (Scharstein & Szeliski, 2002; Mayoral, Lera, & Perez-Ilzarbe, 2006; Hirschmuller & Scharstein, 2007).

Figure 6. General structure of stereo correspondence algorithm



AD is the simplest measure of all. It involves simple subtractions and calculations of absolute values. As a result, it is the most commonly used measure found in literature. The mathematical formulation of AD is:

$$AD(x,y,d) = \left| I_l(x,y) - I_r(x,y-d) \right| \quad (1)$$

where I_l, I_r are the intensity values in left and right image, (x, y) are the pixels coordinates and d is the disparity value under consideration.

SD is somewhat more accurate in expressing the dissimilarity of two pixels. However, the higher computational cost of calculating the square of the intensities' difference is not usually justified by the accuracy gain. It can be calculated as:

$$SD(x,y,d) = \left(I_l(x,y) - I_r(x,y-d) \right)^2 \quad (2)$$

The normalized cross correlation calculates the dissimilarity of image regions instead of single pixels. It produces very robust results, on the cost of computational load. Its mathematical expression is:

$$NCC(x,y,d) = \frac{\sum_{x,y \in W} I_l(x,y) \cdot I_r(x,y-d)}{\sqrt{\sum_{x,y \in W} I_l^2(x,y) \cdot \sum_{x,y \in W} I_r^2(x,y-d)}} \quad (3)$$

where W is the image region under consideration.

The LCDM, as introduced in (Nalpantidis & Gasteratos, 2010b), provides stereo algorithms with tolerance against difficult lighting conditions. The images are initially transformed from the RGB to the HSL colorspace. The transition from the RGB colorspace, which is the usual output of contemporary cameras, to the HSL is straightforward and does not involve any complicated mathematical computations (Gonzalez & Woods, 1992). The HSL colorspace representation is a double cone, as shown in Figure 7(a). In this colorspace, H stands for hue and it determines the human impression about which color (red, green, blue, etc) is depicted. Each color is represented by an angular value ranging between 0 and 360 degrees (0 being red, 120 green and 240 blue). S stands for saturation and determines how vivid or gray the particular color seems. Its value ranges from 0 for gray to 1 for fully saturated (pure) colors. The L channel of the HSL colorspace stands for the Luminosity and it determines the intensity of a specific color. It ranges from 0 for completely dark colors (black) to 1 for fully illuminated colors (white).

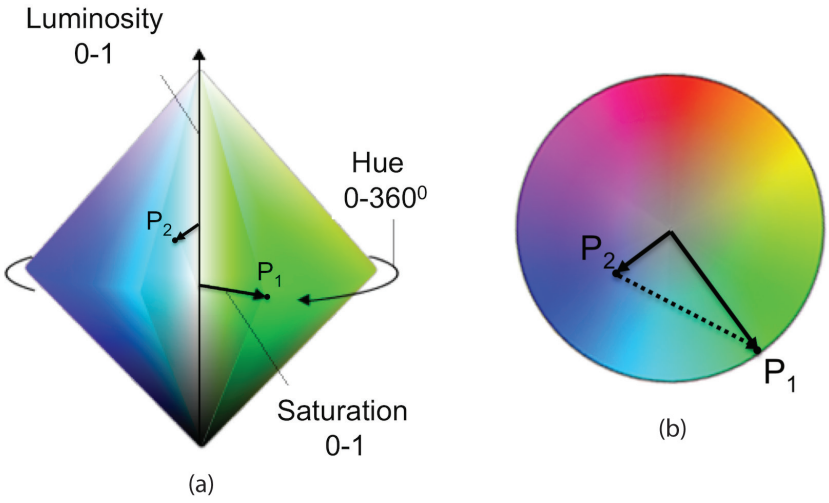
Consequently, the HSL colorspace inherently expresses the lightness of a color and demarcates it from its qualitative characteristics. That is, an object will result in the same values of H and S regardless the environment's illumination conditions. According to this assumption, the proposed dissimilarity measure disregards the values of the L channel in order to calculate the dissimilarity of two colors. The omission of the vertical (L) axis from the colorspace representation leads to 2D circular disk, defined only by H and S, as show in Figure 7(b).

The transition from the 3D colorspace representation to the 2D one, can be conceived as a floor plan projection of the double cone, when observed along the vertical (L) axis. Thus, any color can be described as a planar vector with its initial point being the disc's center. As a consequence, each color P_k can be described as a polar vector or equivalently as a complex number with modulus equal to S_k and argument equal to H_k. That is, a color in the luminosity indifferent colorspace representation can be described as:

$$P_k = S_k e^{iH_k} \quad (4)$$

As a result, the difference, or equivalently the luminosity-compensated dissimilarity measure (LCDM), of two colors P_1 and P_2, shown with dashed line in Figure 7(b) can be calculated as the difference of the two complex numbers:

Figure 7. Views of the HSL color space representation. (a) The double cone representation; and (b) the horizontal slice at L=0



$$LCDM(P_1, P_2) = \left|\vec{P}_1 - \vec{P}_2\right| = \sqrt{S_1^2 + S_2^2 - 2S_1S_2\cos(H_1 - H_2)} \quad (5)$$

This equation is the mathematical formulation of the proposed LCDM dissimilarity measure. It takes into consideration any chromatic information available, except the luminosity. Thus, it can tolerate and compensate for any difference and non-uniformity of the lighting conditions. The proposed measure ignores some information (L), in contrast to the typical AD or SD. As a result, these latter measures are expected to perform somewhat better for totally ideal lighting conditions, which is the case for synthetic and carefully captured test images. On the other hand, any deviation from ideal lighting conditions is supposed to leave the proposed LCDM unaffected, while AD or SD will result in more and more false-matches.

Aggregation Schemes

The dissimilarity values for all the considered disparity values calculated in the first step of a stereo correspondence algorithm comprise the DSI. These results can be aggregated inside fix-sized square windows for constant value of disparity. The width of the window plays an important role on the final result. Small windows generally preserve details but suffer from noise, whereas big windows have the inverse behavior. The window's actual dimensions are chosen so as to keep a balance between the loss of detail and the emergence of noise, given the algorithm's details and the operating situations. The simplest scenario of aggregation is the constant support weight aggregation (CSW), i.e. that of simply summing the values of pixels within each support window.

The summation of the dissimilarity values can also be a weighted one. For instance, in the aggregation scheme used in (Nalpantidis, Sirakoulis, & Gasteratos, 2008a) each pixel is assigned a weight w(i,j,d), the value of which results from the 2D Gaussian function of the pixels Euclidean distance from the central pixel. The center of the function coincides with the central pixel and has a standard deviation equal to the one third of the distance from the central pixel to the nearest window-border. The Gaussian weight function remains the same for fixed width of the support window. Thus, it can be considered as a fixed mask

that can be computed once, and then applied to all the windows.

However, the accuracy of local algorithms that employ CSW aggregation is generally considered low. Indeed, methods that use fixed support regions or even adaptively variable in size and/or shape support regions for aggregation of the computed dissimilarity values have been proven to produce results of inferior quality during the past.

Methods based on extended local methodologies sacrifice some of their computational simplicity in order to obtain more accurate results (Mordohai & Medioni, 2006). Adaptive support weights (ASW) based methods (Yoon & Kweon, 2006a; Gu, Su, Liu, & Zhang, 2008) achieve this, by using fix-sized support windows, whose pixels contribution in the aggregation stage varies depending on their degree of correlation to the windows' central pixel. Despite the acceptance that these methods have enjoyed, the determination of a correlation function is still an active topic.

An ASW-based method for correspondence search is presented in (Yoon & Kweon, 2006a). The support-weights of the pixels in a given support window are adjusted based on color similarity and geometric proximity to reduce the image ambiguity. The difference between pixel colors is measured in the CIELab color space, as this color space is based on measurements on the typical observer and, therefore, the distance of two points in this space is proportional to the stimulus perceived by the human eye. The running time for the Tsukuba image pair with a 35x35 pixels support window is about 0.016 fps on an AMD 2700 processor. The error ratio is 1.29%, 0.97%, 0.99%, and 1.13% for the Tsukuba, Sawtooth, Venus and Map image sets, respectively. These figures can be further improved through a left-right consistency check. The same authors propose a pre-processing step for correspondence search in the presence of specular highlights in (Yoon & Kweon, 2006b). For given input images, specular-free two-band images are generated. The similarity between pixels of these input-image representations can be measured using various correspondence search methods such as the simple SAD-based method, the adaptive support-weights method (Yoon & Kweon, 2006c) and the dynamic programming (DP) method (Lei, Selzer, & Yang, 2006).

Another ASW-based approach is presented in (Nalpantidis & Gasteratos, 2010a). Assigning the right significance weights to each pixel during aggregation has been achieved using the ideas of gestalt theory. The three basic gestalt laws get the following meaning:

- **Proximity (or equivalently Distance):** The closer two pixels are the more correlated to each other they are.
- **Intensity similarity (or equivalently Intensity dissimilarity):** The more similar the colors of two pixels are the more correlated they are.
- **Continuity (or equivalently Discontinuity):** The more similar is the depth of two pixels the more probable it is that they belong to the same larger feature and thus the more correlated they are.

Thus, gestalt theory can be used in order to determine to which degree two pixels are correlated.

The remaining question is exactly how much a correlated pixel to another should contribute to it during the aggregation process. In other words, it is necessary to establish an appropriate mapping between correlation degree and contribution. It is well known, since the 19th century, that HVS interprets physical stimuli in a psychological, non linear rather than in an absolute, linear manner. This psychophysical relationship has been investigated in depth and many explaining theories have been expressed (Pinoli & Debayle, 2007). The Weber-Fechner law is one of those theories and is widely acceptable. It indicates a logarithmic correlation between the subjective perceived intensity and the objective stimulus intensity.

The mathematical expression of this psychophysical law can be derived considering that the

change of perception is proportional to the relative change of the causing stimulus:

$$p = -k \cdot \ln \frac{S}{S_0} \tag{6}$$

where p is the perceived stimulus intensity, S is the actual stimulus intensity, and k is a positive constant determined by the nature of the stimulus. The algorithm presented in (Nalpantidis & Gasteratos, 2010a) calculates the correlation degree by proposing and using a mathematical expression of three gestalt laws, namely the laws of proximity, similarity and continuity. While all the gestalt laws are significant for image understanding applications, these three can be considered to be the essential ones in image processing. Trying to express the gestalt laws in the form of mathematical equations is a difficult task and requires a lot of consideration, as there is no all-satisfying solution. Albeit gestalt psychology theory describes the qualitative characteristics of perceptual organization, a quantitative description, although desired, is not always available. A mathematical expression for the above mentioned gestalt laws has been proposed. The normalized contribution of each of them is subject to the psychophysical law of Weber-Fechner. The mathematical expression is the following:

Let x, y be the coordinates of the central pixel, x', y' the coordinates of a pixel lying inside its support region and d the disparity value currently being considered. Proximity of the two pixels is taken into consideration using their Euclidean distance on the image plain. The distance of the pixel (x',y') from the pixel (x, y) is calculated as:

$$distance(x',y')_{(x,y)} = \sqrt{(x-x')^2 - (y-y')^2} \tag{7}$$

The color dissimilarity of the two pixels can be estimated by the AD of their color intensities. This metric should not be confused with the AD calculated in the first step of the algorithm, since the former ones are calculated for pixels of the same image. Thus, the dissimilarity between the pixels (x', y') and (x, y) is calculated as:

$$dissimilarity(x',y')_{(x,y)} = \frac{1}{3}\sum_{C\in R,G,B}\left|I_C(x,y) - I_C(x',y')\right| \tag{8}$$

However, the AD calculated in the first step of this algorithm can be used to estimate the continuity of the pixels (x, y) and (x', y'). The continuity of two pixels can be described by the possibility that they both have the same depth, i.e. to share the same disparity value. The normalization of the AD calculated in the first step, results in an expression of the possibility that the true disparity value for the pixel (x', y') is not d. This possibility measure express the complement of continuity, i.e. the discontinuity. The less likely it is for a pixel (x', y') to have a disparity value d, the less it should bias the central pixel (x, y) in favor of the same disparity value d. The discontinuity between the pixels (x', y') and (x, y) is calculated as:

$$discontinuity(x',y',d)_{(x,y,d)} = \frac{AD(x',y',d)}{\max(AD)} \tag{9}$$

The last three equations quantify the gestalt theory. On the other hand the exact impact of those expression on the final result, is obtained by applying the Weber-Fechner. The values for distance, dissimilarity and discontinuity used hereafter are normalized towards its respective maximum value. Consequently, the factor S_o of the Weber-Fechner law for this case is equal to one and can be neglected. Thus, the weighting factor due to each gestalt law can be calculated:

$$w_{dist}(x',y',d)_{(x,y,d)} = -k_1 \cdot \ln\left(distance(x',y',d)_{(x,y,d)}\right) \tag{10}$$

$$w_{dissim}(x',y',d)_{(x,y,d)} = -k_2 \cdot \ln\left(dissimilarity(x',y',d)_{(x,y,d)}\right) \tag{11}$$

$$w_{discon}(x',y',d)_{(x,y,d)} = -k_3 \cdot \ln\left(discontinuity(x',y',d)_{(x,y,d)}\right) \tag{12}$$

These three weights are combined into one by multiplication, providing a general total weight:

$$w_{tot} = w_{dist} \cdot w_{dissim} \cdot w_{discon} \tag{13}$$

The total weight is calculated for both the left and the right input images, obtaining $w_{tot,l}$ and $w_{tot,r}$, respectively. However, distance and discontinuity are the same for both images considering the same pixel. Consequently, only dissimilarity has to be separately calculated for each image. Finally, the ASW aggregation, taking into consideration the weighting factor for each pixel is performed and results into the aggregated DSI:

$$DSI(x,y,d) = \frac{\sum\left(w_{tot,l} \cdot w_{tot,r} \cdot AD(x,y,d)\right)}{\sum\left(w_{tot,l} \cdot w_{tot,r}\right)} \tag{14}$$

TRAVERSABILITY ESTIMATION

Autonomous robots' behavior greatly depends on the accuracy of their decision-making algorithms. Reliable depth estimation is commonly needed in numerous autonomous behaviors. Autonomous navigation (Hariyama, Takeuchi, & Kameyama, 2000), obstacle avoidance (Nalpantidis, Kostavelis, & Gasteratos, 2009), localization and mapping, and traversability estimation are just a few of them (Murray & Little, 2000; Sim & Little, 2009). Vision-based solutions are becoming more and more attractive due to their decreasing cost as well as their inherent coherence with human imposed mechanisms. In the case of stereo vision-based navigation, the accuracy and the refresh rate of the computed disparity maps are the cornerstone of its success (Iocchi & Konolige, 1998; Schreer, 1998). However, robotic applications place strict requirements on the demanded speed and accuracy of vision depth-computing algorithms. Depth estimation using stereo vision, comprises the stereo correspondence problem. Stereo correspondence is known to be very computational demanding. The computation of dense and accurate depth images, i.e. disparity maps, in frame rates suitable for robotic applications is an open problem for the scientific community. Most of the attempts to confront the demand for accuracy focus on the development of sophisticated stereo correspondence algorithms, which usually increase the computational load exponentially. On the other hand, the need for real-time frame rates, inevitably, imposes compromises concerning the quality of the results. However, results' reliability is of crucial importance for autonomous robotic applications.

A wide range of sensors and various methods have been proposed in the relevant literature, as far as traversability estimation techniques are concerned. Some interesting details about the developed sensor systems and proposed detection and avoidance algorithms can be found in (Borenstein & Koren, 1990) and (Ohya, Kosaka, & Kak, 1998). Movarec has proposed the Certainty Grid method in (Moravec, 1987) and Borenstein (Borenstein & Koren, 1991) has proposed the Virtual Force Field method for robot obstacle avoidance. Then the Elastic Strips method was proposed in (Khatib, 1996, 1999) treating the trajectory of the robot as an elastic material to avoid obstacles. Moreover, (Kyung Hyun, Minh Ngoc, & M. Asif Ali, 2008) present a modified Elastic Strip method for mobile robots operating in uncertain environments. Review of popular obstacle avoidance algorithms covering them in more detail can be found in (Manz, Liscano, & Green, 1993) and (Kunchev, Jain, Ivancevic, & Finn, 2006).

The traversability estimation systems found in literature involve the use of one or a combination of ultrasonic, Laser. infrared (IR) or vision senors (Siegwart & Nourbakhsh, 2004). The use of ultrasonic, Laser and IR sensors is well-studied and the depth measurements are quite accurate and easily available. However, such sensors suffer either from achieving only low refresh rates (Vandorpe, Van Brussel, & Xu, 1996) or being extremely expensive. On the other hand vision sensors, either monocular, stereo or multicamera ones, can combine high frame rates and appealing prices.

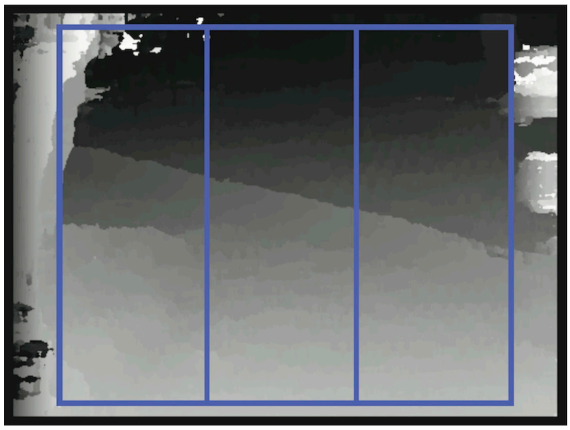
Stereo vision is often used in vision-based methods, instead of monocular sensors, due to the simpler calculations involved in the depth estimation. Regarding stereo vision systems, one of the most popular methods for obstacle avoidance is the initial estimation of the so called v-disparity image (De Cubber et al., 2009). This method requires complex calculations and is applied in order to confront the noise in low quality disparity images (Labayrade et al., 2002; Zhao et al., 2007; Soquet et al., 2007). However, if detailed and noise-free disparity maps were available, less complicated methods could have been used.

Such a method is found in (Nalpantidis, Kostavelis, & Gasteratos, 2009). The disparity map obtained by a stereo correspondence algorithm is used to extract useful information about the navigation of a robot. Contrary to many implementations that involve complex calculations upon the disparity map, the proposed decision making algorithm involves only simple summations and checks. This is feasible due to the absence of significant noise in the produced disparity map. The goal of the developed algorithm is to detect any existing obstacles in front of the robot and to safely avoid it, by steering the robot left, right or to moving it forward. In order to achieve that, the developed method divides the disparity map into three windows, as in Figure 8.

In the central window, the pixels p whose disparity value $D(p)$ is greater than a defined threshold value T are enumerated. Then, the enumeration result is examined. If it is smaller than a predefined rate r of all the central windows pixels, this means that there are no obstacles detected exactly in front of the robot and in close distance, and thus the robot can move forward. On the other hand, if this enumeration's result exceeds the predefined rate, the algorithm examines the other two windows and chooses the one with the smaller average disparity value. In this way the window with the fewer obstacles will be selected. The values of the parameters T and r play an important role to the algorithm's behavior. Small values of T in conjunction with small values of r favor the hesitancy in moving forward, ensuring obstacle avoidance but at the same time being susceptible to false alarms due to noise. On the other hand, the opposite scenario is less susceptible to false alarms but may be proven risky.

Traversability estimation is also a significant part of visual mapping applications. A 2D map can be computed from stereo image pairs. Using the disparity map obtained form a stereo correspondence algorithm a reliable v-disparity image can be computed (Labayrade et al., 2002; Zhao et al., 2007). The terrain in the v-disparity image is modeled by a linear equation. The parameters of this linear equation can be found using Hough transform (De Cubber et al., 2009), if the camera-

Figure 8. Depth map's division in three windows



environment system's geometry is unknown. However, if the geometry of the system is constant and known (which is the case for a camera firmly mounted on a robot exploring a flat, e.g. indoor, environment) the two parameters can be easily computed beforehand and used in all the image pairs during the exploration. A tolerance region on either side of the terrain's linear segment is considered and any point outside this region is considered as an "obstacle". The linear segments denoting the terrain and the tolerance region overlaid on the v-disparity image are shown in Figure 9(a). For each pixel corresponding to an "obstacle" the local coordinates are computed. The local map, e.g. the one shown in Figure 9(b), is an occupancy grid of the environment consisting of all the points corresponding to "obstacles".

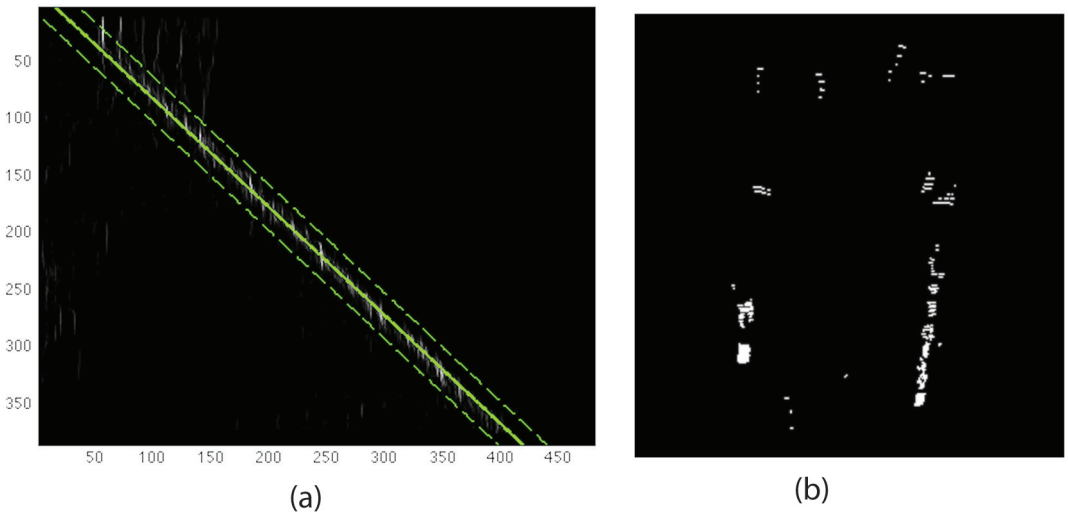
CONCLUSION

Stereo vision is a tested, useful and popular tool for inferring the depth of a scene with only passive optical sensors. Robotics, on the other hand, evolves rapidly and demand methods that can serve autonomous behaviors. Within this context, stereo correspondence algorithms need to provide accurate depth maps, in real-time frame-rates, confronting, at the same time, any difficulties imposed by the robots' environments.

In this chapter, the most interesting research issues of the robotics-oriented stereo vision field have been covered and solutions and possibilities have been presented. Such issues involve the handling of non-ideal lighting conditions, the requirement for simple calculation schemes, the use of multi-view stereo systems, the handling of miscalibrated image sensors, and the introduction of new biologically inspired methods to robotic vision. Various stereo correspondence algorithms that have non-iterative computational structure and are able to cope with real life images have been discussed. The dissimilarity measures, as well as the aggregation schemes that they employ have been examined.

Since many stereo vision-based robotic applications demand such characteristics, the presented stereo correspondence algorithms comprise effective solutions, which can be used as the cornerstone of more advanced autonomous robotic behaviors. Last, such applications of stereo vision within the domain of mobile robotic applications are covered. More specifically, the use of the obtained depth maps by algorithms that analyze the traversability

Figure 9. (a) V-disparity images for the image and (b) the corresponding local map



of the field in order the robot to avoid possible obstacles has been examined.

REFERENCES

Agrawal, M., Konolige, K., & Bolles, R. (2007). *Localization and mapping for autonomous navigation in outdoor terrains: A stereo vision approach*. In IEEE Workshop on Applications of Computer Vision. Austin, Texas, USA.

Barnard, S. T., & Thompson, W. B. (1980). Disparity analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *2*(4), 333–340. doi:10.1109/TPAMI.1980.4767032

Berthouze, L., & Metta, G. (2005). Epigenetic robotics: Modelling cognitive development in robotic systems. *Cognitive Systems Research*, *6*(3), 189–192. doi:10.1016/j.cogsys.2004.11.002

Binaghi, E., Gallo, I., Marino, G., & Raspanti, M. (2004). Neural adaptive stereo matching. *Pattern Recognition Letters*, *25*(15), 1743–1758. doi:10.1016/j.patrec.2004.07.001

Borenstein, J., & Koren, Y. (1990). Real-time obstacle avoidance for fast mobile robots in cluttered environments. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(5), 1179–1187. doi:10.1109/21.44033

Borenstein, J., & Koren, Y. (1991). The vector field histogram-fast obstacle avoidance for mobile robot. *IEEE Transactions on Robotics and Automation*, *7*(3), 278–288. doi:10.1109/70.88137

Brockers, R., Hund, M., & Mertsching, B. (2005). Stereo vision using cost-relaxation with 3D support regions. In *Image and Vision Computing New Zealand* (pp. 96-101).

Corke, P. (2005, November). Machine vision toolbox. *IEEE Robotics & Automation Magazine*, *12*(4), 16–25. doi:10.1109/MRA.2005.1577021

De Cubber, G., Doroftei, D., Nalpantidis, L., Sirakoulis, G. C., & Gasteratos, A. (2009). *Stereobased terrain traversability analysis for robot navigation*. In IARP/EURON workshop on robotics for risky interventions and environmental surveillance. Brussels, Belgium.

Forsyth, D. A., & Ponce, J. (2002). *Computer vision: A modern approach*. Upper Saddle River, NJ: Prentice Hall.

Gonzalez, R. C., & Woods, R. E. (1992). *Digital image processing*. Boston, MA: AddisonWesley Longman Publishing Co., Inc.

Gu, Z., Su, X., Liu, Y., & Zhang, Q. (2008). Local stereo matching with adaptive support-weight, rank transform and disparity calibration. *Pattern Recognition Letters*, *29*(9), 1230–1235. doi:10.1016/j.patrec.2008.01.032

Hariyama, M., Takeuchi, T., & Kameyama, M. (2000). Reliable stereo matching for highly-safe intelligent vehicles and its VLSI implementation. In *IEEE Intelligent Vehicles Symposium* (p. 128133).

Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision* (2nd ed.). Cambridge University Press. doi:10.1017/CBO9780511811685

Hirschmuller, H., & Scharstein, D. (2007, June). *Evaluation of cost functions for stereo matching*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Minneapolis, Minnesota, USA.

Hogue, A., German, A., & Jenkin, M. (2007). Underwater environment reconstruction using stereo and inertial data. In *IEEE International Conference on Systems, Man and Cybernetics* (p. 2372-2377). Montreal, Canada.

Iocchi, L., & Konolige, K. (1998). *A multiresolution stereo vision system for mobile robots*. In Italian AI Association Workshop on New Trends in Robotics Research.

Jeong, H., & Park, S. (2004). Generalized trellis stereo matching with systolic array. In *International Symposium on Parallel and Distributed Processing and Applications* (vol. 3358, p. 263-267). Springer Verlag.

Kelly, A., & Stentz, A. (1998, May). *Stereo vision enhancements for low-cost outdoor autonomous vehicles*. In International Conference on Robotics and Automation, Workshop Ws-7, Navigation of Outdoor Autonomous Vehicles.

Khatib, O. (1996). Motion coordination and reactive control of autonomous multi-manipulator system. *Journal of Robotic Systems*, *15*(4), 300–319.

Khatib, O. (1999). Robot in human environments: Basic autonomous capabilities. *The International Journal of Robotics Research*, *18*(7), 684–696. doi:10.1177/02783649922066501

Klancar, G., Kristan, M., & Karba, R. (2004). Wide-angle camera distortions and non-uniform illumination in mobile robot tracking. *Journal of Robotics and Autonomous Systems*, *46*, 125–133. doi:10.1016/j.robot.2003.11.001

Konolige, K., Agrawal, M., Bolles, R. C., Cowan, C., Fischler, M., & Gerkey, B. P. (2006). Outdoor mapping and navigation using stereo vision. In *International Symposium on Experimental Robotics* (vol. 39, pp. 179-190). Brazil: Springer.

Kunchev, V., Jain, L., Ivancevic, V., & Finn, A. (2006). Path planning and obstacle avoidance for autonomous mobile robots: A review. In International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (vol. 4252, pp. 537-544). Springer-Verlag.

Kyung Hyun, C., Minh Ngoc, N., & Asif Ali, R. (2008). A real time collision avoidance algorithm for mobile robot based on elastic force. *International Journal of Mechanical . Industrial and Aerospace Engineering*, *2*(4), 230–233.

Labayrade, R., Aubert, D., & Tarel, J.-P. (2002). Real time obstacle detection in stereovision on non flat road geometry through V-disparity representation. In *IEEE Intelligent Vehicle Symposium* (vol. 2, pp. 646-651). Versailles, France.

Lei, C., Selzer, J., & Yang, Y.-H. (2006). Region-tree based stereo using dynamic programming optimization. *IEEE Conference on Computer Vision and Pattern Recognition, 2,* 2378-2385.

Manz, A., Liscano, R., & Green, D. (1993). A comparison of realtime obstacle avoidance methods for mobile robots . In *Experimental Robotics ii* (pp. 299–316). Springer-Verlag. doi:10.1007/BFb0036147

Masrani, D. K., & MacLean, W. J. (2006). A real-time large disparity range stereo-system using FPGAS. In *IEEE International Conference on Computer Vision Systems* (vol. 3852, p. 13-20).

Mayoral, R., Lera, G., & Perez-Ilzarbe, M. J. (2006). Evaluation of correspondence errors for stereo. *Image and Vision Computing*, *24*(12), 1288–1300. doi:10.1016/j.imavis.2006.04.006

Mead, C. (1990). Neuromorphic electronic systems. *Proceedings of the IEEE*, *78*(10), 1629–1636. doi:10.1109/5.58356

Mingxiang, L., & Yunde, J. (2006, Dec). Trinocular cooperative stereo vision and occlusion detection. IEEE International Conference on Robotics and Biomimetics, 1129-1133.

Moravec, P. (1987). Certainty grids for mobile robots. In *NASA/JPL Space Telerobotics Workshop* (Vol. 3, pp. 307-312).

Mordohai, P., & Medioni, G. G. (2006). Stereo using monocular cues within the tensor voting framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(6), 968982. doi:10.1109/TPAMI.2006.129

Muhlmann, K., Maier, D., Hesser, J., & Manner, R. (2002). Calculating dense disparity maps from color stereo images, an efficient implementation. *International Journal of Computer Vision*, *47*(1-3), 79–88. doi:10.1023/A:1014581421794

Murray, D., & Little, J. J. (2000). Using real-time stereo vision for mobile robot navigation. *Autonomous Robots*, *8*(2), 161–171. doi:10.1023/A:1008987612352

Nalpantidis, L., Amanatiadis, A., Sirakoulis, G. C., & Gasteratos, A. (in press). An efficient hierarchical matching algorithm for processing uncalibrated stereo vision images and its hardware architecture. *IET Image Processing*.

Nalpantidis, L., Chrysostomou, D., & Gasteratos, A. (2009, December). Obtaining reliable depth maps for robotic applications with a quad-camera system. In International Conference *on Intelligent Robotics and Applications* (vol. 5928, p. 906-916). Singapore: Springer-Verlag.

Nalpantidis, L., & Gasteratos, A. (2010a). Biologically and psychophysically inspired adaptive support weights algorithm for stereo correspondence. *Robotics and Autonomous Systems*, *58*, 457–464. doi:10.1016/j.robot.2010.02.002

Nalpantidis, L., & Gasteratos, A. (2010b). Stereo vision for robotic applications in the presence of non-ideal lighting conditions. *Image and Vision Computing*, *28*, 940–951. doi:10.1016/j.imavis.2009.11.011

Nalpantidis, L., Kostavelis, I., & Gasteratos, A. (2009). Stereovision-based algorithm for obstacle avoidance. In *International Conference on Intelligent Robotics and Applications* (vol. 5928, pp. 195-204). Singapore: Springer-Verlag.

Nalpantidis, L., Sirakoulis, G. C., & Gasteratos, A. (2008a). A dense stereo correspondence algorithm for hardware implementation with enhanced disparity selection. In *5th Hellenic Conference on Artificial Intelligence* (vol. 5138, pp. 365-370). Syros, Greece: Springer-Verlag.

Nalpantidis, L., Sirakoulis, G. C., & Gasteratos, A. (2008b). Review of stereo vision algorithms: From software to hardware. *International Journal of Optomechatronics*, *2*(4), 435–462. doi:10.1080/15599610802438680

Ogale, A. S., & Aloimonos, Y. (2005a, April). Robust contrast invariant stereo correspondence. In *IEEE International Conference on Robotics and Automation* (pp. 819-824).

Ogale, A. S., & Aloimonos, Y. (2005b). Shape and the stereo correspondence problem. *International Journal of Computer Vision*, *65*(3), 147–162. doi:10.1007/s11263-005-3672-3

Ogale, A. S., & Aloimonos, Y. (2007). 04). A roadmap to the integration of early visual modules. *International Journal of Computer Vision*, *72*(1), 9–25. doi:10.1007/s11263-006-8890-9

Ohya, A., Kosaka, A., & Kak, A. (1998). Vision-based navigation of mobile robot with obstacle avoidance by single camera vision and ultrasonic sensing. *IEEE Transactions on Robotics and Automation*, *14*(6), 969–978. doi:10.1109/70.736780

Park, S., & Jeong, H. (2007). Real-time stereo vision FPGA chip with low error rate. In *International Conference on Multimedia and Ubiquitous Engineering* (pp. 751-756).

Pinoli, J. C., & Debayle, J. (2007). Logarithmic adaptive neighborhood image processing (LANIP): Introduction, connections to human brightness perception, and application issues. *EURASIP Journal on Advances in Signal Processing*, (1): 114–135.

Ruigang, Y., Welch, G., & Bishop, G. (2002). Real-time consensus-based scene reconstruction using commodity graphics hardware. *10th Pacific Conference on Computer Graphics and Applications*, (pp. 225-234).

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, *47*(1-3), 7–42. doi:10.1023/A:1014573219977

Schirmacher, H., Li, M., & Seidel, H.-P. (2001). *On-the-fly processing of generalized lumigraphs* (pp. 165–173). Eurographics.

Schreer, O. (1998). Stereo vision-based navigation in unknown indoor environment. In *5th European Conference on Computer Vision* (vol. 1, pp. 203-217).

Shimonomura, K., Kushima, T., & Yagi, T. (2008). Binocular robot vision emulating disparity computation in the primary visual cortex. *Neural Networks*, *21*(2-3), 331–340. doi:10.1016/j.neunet.2007.12.033

Siegwart, R., & Nourbakhsh, I. R. (2004). *Introduction to autonomous mobile robots*. Massachusetts: MIT Press.

Sim, R., & Little, J. J. (2009). Autonomous vision-based robotic exploration and mapping using hybrid maps and particle filters. *Image and Vision Computing, 27*(1-2), 167-177. (Canadian Robotic Vision 2005 and 2006)

Soquet, N., Aubert, D., & Hautiere, N. (2007). Road segmentation supervised by an extended V-disparity algorithm for autonomous navigation. In *IEEE Intelligent Vehicles Symposium* (pp. 160-165). Istanbul, Turkey.

Sun, C., & Peleg, S. (2003). Fast panoramic stereo matching using cylindrical maximum surfaces. IEEE Trans. *SMC Part B*, *34*, 760–765.

Vandorpe, J., Van Brussel, H., & Xu, H. (1996). Exact dynamic map building for a mobile robot using geometrical primitives produced by a 2d range finder. In IEEE International Conference on Robotics and Automation (pp. 901-908). Minneapolis, USA.

Wiegand, T., Sullivan, G., Bjntegaard, G., & Luthra, A. (2003). Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(7), 560–576. doi:10.1109/TCSVT.2003.815165

Wilburn, B., Smulski, M., Lee, K., & Horowitz, M. A. (2002). The light field video camera. In *Media Processors* (p. 29-36).

Yang, J. C., Everett, M., Buehler, C., & Mcmillan, L. (2002). A real-time distributed light field camera. In *Eurographics Workshop on Rendering* (pp. 77-86).

Yin, P., Tourapis, H., Tourapis, A., & Boyce, J. (2003). Fast mode decision and motion estimation for JVT/H.264. In *IEEE International Conference on Image Processing* (vol. 3, pp. 853–856).

Yoon, K.-J., & Kweon, I. S. (2006a). Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(4), 650–656. doi:10.1109/TPAMI.2006.70

Yoon, K.-J., & Kweon, I. S. (2006b). Correspondence search in the presence of specular highlights using specular-free two-band images. In *7th Asian Conference on Computer Vision* (vol. 3852, pp. 761-770). Hyderabad, India: Springer.

Yoon, K.-J., & Kweon, I. S. (2006c). Stereo matching with symmetric cost functions. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (vol. 2, pp. 2371-2377).

Zach, C., Karner, K., & Bischof, H. (2004). Hierarchical disparity estimation with programmable 3D hardware. In *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision* (pp. 275-282).

Zhao, J., Katupitiya, J., & Ward, J. (2007). Global correlation based ground plane estimation using V-disparity image. In *IEEE International Conference on Robotics and Automation* (pp. 529-534). Rome, Italy.

Zitnick, C. L., & Kanade, T. (2000). A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(7), 675–684. doi:10.1109/34.865184

KEY TERMS AND DEFINITIONS

Dense Stereo Correspondence Algorithm: A stereo correspondence algorithm that estimates disparity values for all the image pixels.

Disparity Map: An image constituted by the disparity values of each pixel, being thus equivalent to a depth map.

Disparity: The difference of an observed point's image coordinates when viewed under different viewpoints.

Dissimilarity Measure: A function that quantitatively expresses how much dissimilar two image pixels are.

Sparse Stereo Correspondence Algorithm: A stereo correspondence algorithm that estimates disparity values for some of the image pixels.

Stereo Correspondence: The procedure of matching pixels between two images that derive from the same scene.

Traversability Estimation: The procedure of determining whether there are obstacles or not in a field.