AN INTELLIGENT MULTI SENSOR SYSTEM FOR A HUMAN ACTIVITIES SPACE

ASPECTS OF QUALITY MEASUREMENT AND SENSOR ARRANGEMENT

Jiandan Chen

Blekinge Institute of Technology
Doctoral Dissertation Series No. 2011:05

School of Engineering

An Intelligent Multi Sensor System for a Human Activities Space

Aspects of Quality Measurement and Sensor Arrangement

Jiandan Chen

Blekinge Institute of Technology Doctoral Dissertation Series
No 2011:05

An Intelligent Multi Sensor System for a Human Activities Space

Aspects of Quality Measurement and Sensor Arrangement

Jiandan Chen

Department of Electrical Engineering
School of Engineering
Blekinge Institute of Technology
SWEDEN

Department of Electrical Engineering
School of Engineering
Publisher: Blekinge Institute of Technology
Printed by Printfabriken, Karlskrona, Sweden 2011
ISBN: 978-91-7295-200-3
Blekinge Institute of Technology Doctoral Dissertation Series
ISSN 1653-2090
urn:nbn:se:bth-00487

Abstract

In our society with its aging population, the design and implementation of a high-performance distributed multi-sensor and information system for autonomous physical services become more and more important. In line with this, this thesis proposes an Intelligent Multi-Sensor System, IMSS, that surveys a human activities space to detect and identify a target for a specific service. The subject of this thesis covers three main aspects related to the set-up of an IMSS: an improved depth measurement and reconstruction method and its related uncertainty, a surveillance and tracking algorithm and finally a way to validate and evaluate the proposed methods and algorithms.

The thesis discusses how a model of the depth spatial quantisation uncertainty can be implemented to optimize the configuration of a sensor system to capture information of the target objects and their environment with required specifications. The thesis introduces the dithering algorithm which significantly reduces the depth reconstruction uncertainty. Furthermore, the dithering algorithm is implemented on a sensor-shifted stereo camera, thus simplifying depth reconstruction without compromising the common stereo field of view.

To track multiple targets continuously, the Gaussian Mixture Probability Hypothesis Density, GM-PHD, algorithm is implemented with the help of vision and Radio Frequency Identification, RFID, technologies. The performance of the tracking algorithm in a vision system is evaluated by a circular motion test signal. The thesis introduces constraints to the target space, the stereo pair characteristics and the depth reconstruction accuracy to optimize the vision system and to control the performance of surveillance and 3D reconstruction through integer linear programming. The human being within the activity space is modelled as a tetrahedron, and a field of view in spherical coordinates are used in the control algorithms.

In order to integrate human behaviour and perception into a technical system, the proposed adaptive measurement method makes use of the Fuzzily Defined Variable, FDV. The FDV approach enables an estimation of the quality index based on qualitative and quantitative factors for image quality evaluation using a neural network.

The thesis consists of two parts, where Part I gives an overview of the applied theory and research methods used, and Part II comprises the eight papers included in the thesis.

***Keywords**: 3D Reconstruction, Depth Measurement, Depth Reconstruction, Dither, Human Factor, Image Quality, Iso-disparity Surfaces, Multi-Sensor System, Quality Measurement, Sensor Arrangement, Surveillance, Tracking, Uncertainty.*

Acknowledgements

First of all, I would like to express my sincere gratitude to my examiner and main supervisor Prof. Wlodek Kulesza for giving me the opportunity to conduct the research that I love at the Blekinge Institute of Technology, BTH. He is a great mentor and has given me important advice and crucial guidance during my research work, and he possesses profound knowledge within the fields of measurement science, research methodology and signal processing. I am sincerely grateful for his untiring work and for the countless hours he has devoted to our papers and to my thesis. His efforts and contribution made this thesis possible. I am also most thankful to my co-supervisor, Dr. Siamak Khatibi for his profound knowledge and experience in the field of computer vision and image processing, and I wish to extend my sincere gratitude to him for engaging in many fruitful discussions and providing numerous creative ideas. Furthermore, I am very grateful to my co-supervisor Dr. Benny Lövström who has supported and helped me throughout the whole research project and want to express my appreciation for his discussions and comments on the papers and this thesis. He has also helped me enjoy life at BTH. Last, but definitely not least, I want to thank Prof. Ingvar Claesson for examining my licentiate dissertation.

I would also like to thank former and present colleagues at the Blekinge Institute of Technology and former colleagues at the University of Kalmar for being so helpful, friendly and cheerful. The people working at these departments made me feel at home and there is always such a pleasant atmosphere at the BTH School of Engineering.

I want to take this opportunity to thank my former colleague Dr. Jenny Wirandi for fruitful discussions and encouragement. I wish to acknowledge Prof. Stefan Andersson-Engels, Pontus Svenmarker and Dr. Zuguang Guan at Lund University, and Dr. Fredrik Bergholm at the Royal Institute of Technology for lending their expertise regarding the necessary laboratory equipment. I also wish to thank Fredrik for coming to Ronneby for a fruitful discussion about the project and Dr. Anders M. Johansson for providing knowledge and inspiration on tracking during his course. I am also very grateful to Dr. Johan Höglund at Akademiska Språkbyrån, and to Paul Curley for their comments.

My appreciation goes as well to the Master students Wail Mustafa, Abu Bakr Siddig, Soheil Ghadami, Oyekanlu Emmanuel Adebomi, Onidare Samuel Olusayo, Iyeyinka Damilola Olayanju and Olabode Paul Ojelabi for their cooperation and discussions.

I would like to acknowledge and thank the Education Section at the Embassy of China in Sweden and the China Scholarship Council for nominating me to the Chinese Government Award for Outstanding Self-Financed Students Abroad.

Finally, I am forever grateful to my wife, Haiyan, for her patience, endless love and support for my work. She moved with me to Sweden, she always stands by me, and she has given our life together so much happiness. Very special thanks to my daughter Zihan who gives me the inspiration and the energy I need to keep working. I also want to thank the rest of my family for their love and support.

Karlskrona, January 2011
Jiandan Chen

Contents

Acronyms

ADC	Analogue to Digital Converter
AF	Accuracy Factor
AGP	Art Gallery Problem
CCD	Charge Coupled Device
FDV	Fuzzily Defined Variable
FoV	Field of View
FISST	Finite Set Statistics
GM-PHD	Gaussian Mixture Probability Hypothesis Density
GPS	Global Positioning System
GUM	Guide to the Expression of Uncertainty in Measurement
ILP	Integer Linear Programming
IMSS	Intelligent Multi-Sensor System
IR	Image Resolution
IVAS	Intelligent Vision Agent System
JPEG	Joint Photographic Experts Group
JPDAF	Joint Probabilistic Data Association Filter
LM	Levenberg-Marquardt
MHT	Multiple Hypothesis Tracking
MIT	Massachusetts Institute of Technology
NN	Neural Network
OSPA	Optimal Sub-Pattern Assignment
PCA	Principal Components Analysis
PDF	Probability Distribution Function
PHD	Probability Hypothesis Density
PPM	Perspective Projection Matrix
QDA	Quantitative Descriptive Analysis
QI	Quality Indices
QIM	Quality Index Method
RFID	Radio Frequency Identification
RFS	Random Finite Sets
RSS	Radio Signal Strength
RMSD	Root Mean Square Deviation
WQI	Water Quality Index

List of Appended Papers

This thesis is based on the following papers. In the text, they are referred to by their Roman numerals according to their logical order as stated below:

I. J. Chen, S. Khatibi, J. Wirandi, and W. Kulesza, "Planning of a multiple sensor system for a human activities space – aspects of iso-disparity surface", *Proceedings of SPIE on Optics and Photonics in Security and Defence*, vol. 6739, Florence, Italy, September, 2007.

II. J. Chen, S. Khatibi, and W. Kulesza, "Depth reconstruction uncertainty analysis and improvement – the dithering approach", *Elsevier Journal of Image and Vision Computing*, vol. 28, no. 9, pp. 1377-1385, September, 2010.

III. J. Chen, W. Mustafa, A. Siddig, and W. Kulesza, "Applying dithering to improve depth measurement using a sensor-shifted stereo camera", *Metrology and Measurement Systems*, vol. 17, no. 3, pp. 335-348, October, 2010.

IV. J. Chen, D. I. Olayanju, O. P. Ojelabi, and W. Kulesza, "RFID multi-target tracking using the probability hypothesis density algorithm for a health care application", *the 3rd International ICST Conference on IT Revolutions*, Córdoba, Spain, March, 2011 (in Print).

V. J. Chen, S. Khatibi, and W. Kulesza, "Planning of a multi stereo visual sensor system for a human activities space", *Proceedings of the 2nd International Conference on Computer Vision Theory and Applications*, pp. 480-485, Barcelona, Spain, March, 2007.

VI. J. Chen, S. Khatibi, and W. Kulesza, "Planning of a multi stereo visual sensor system - depth accuracy and variable baseline approach", *Proceedings of IEEE Computer Society 3DTV-Con, the True Vision Capture, Transmission and Display of 3D Video*, Kos, Greece, May, 2007.

VII. J. Chen, S. Ghadami, and W. Kulesza, "Evaluation of the GM-PHD filter for multi-target tracking with a stereo vision system", *IEEE International Instrumentation and Measurement Technology Conference*, Hangzhou, China, May, 2011 (in Print).

VIII. J. Wirandi, J. Chen, and W. Kulesza, "An adaptive quality assessment system – aspect of human factor and measurement uncertainty", *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 11, pp. 68-75, January, 2009 (Recipient of the Andi Chi Best Paper Award by the IEEE Instrumentation and Measurement Society, 2010).

Other publications not included in this thesis, but produced during my doctoral programme and related to the subject of the thesis:

IX. J. Chen, O. E. Adebomi, O. S. Olusayo, and W. Kulesza, "The evaluation of the Gaussian mixture probability hypothesis density approach for multi-target tracking", *IEEE International Conference on Imaging Systems and Techniques*, Thessaloniki, Greece, July 2010.

X. J. Chen, "The depth reconstruction accuracy in a stereo vision system", *Metrologia:dzis i jutro*, pp. 123-132, September, 2009, ISBN: 83-911669-5-3 (Recipient of the Best Lecture Award at the 41st Metrology Conference, Gdansk, Poland, 2009).

XI. W. Kulesza, J. Chen, and S. Khatibi, "Arrangement of a multi stereo visual sensor system for a human activities space", in: A. Bhatti (Ed.), *Stereo Vision*, pp. 153-172, InTech Education and Publishing, Vienna, Austria, November, 2008, ISBN: 978-953-7619-22-0.

XII. J. Chen, "A multi sensor system for a human activities space – aspects of planning and quality measurement", *Licentiate Dissertation*, Blekinge Institute of Technology, Karlskrona, Sweden, August, 2008, ISBN: 978-91-7295-147-1.

XIII. J. Wirandi, J. Chen, and W. Kulesza, "An adaptive model of the fuzzy variable – quality index", *AMUEM 2007 – International IEEE Workshop on Advanced Methods for Uncertainty Estimation in Measurement*, Trento, Italy, July, 2007.

XIV. J. Chen, S. Khatibi, B. Lövström, and W. Kulesza, "The rectification effect on the depth measure uncertainty – a case study of a calibrated stereo vision system", *EURASIP Journal on Image and Video Processing*, 2011 (Manuscript).

Part I

1 Introduction

1.1 Background

As the size of the elderly population increases, healthcare services are increasingly unable to cope, and are therefore searching for a new public health paradigm. Autonomous physical services that support and take care of elderly people by doing housework and providing a comfortable living environment are becoming more and more in demand in our society. For this reason, it is of great importance to conduct research towards the design and implementation of a high-performance autonomous distributed multi-sensor information system which can understand human behaviour and living environments, and thereby enhance living services and quality.

"Bringing abundant computation and communication, as pervasive and free as air, naturally into people's lives" is proposed by the MIT Oxygen Project, [1]. The human-centred computation is focused on human needs and abilities instead of on the needs and possibilities of the machine. Furthermore, Hashimoto has suggested the concept of *intelligent space* which can be defined as "space with functions that can provide appropriate services for human beings by capturing events in the space and by utilizing the information intelligently with computers and robots", [2]. The intelligent space is thus a space that can be treated as a platform which supports people both informationally and physically. In this way, the intelligent space is an interface both for the human being and for robots.

The proposed Intelligent Multi-Sensor System, IMSS, is a high-performance autonomous distributed vision and information processing system. Figure 1.1 shows that the system consists of multiple sensors and actuators for surveillance of a human activities space which includes the human being and the surrounding environment, including robots, household appliances, lights, and so on. The system does not only gather information, but also controls the vision sensors including their deployment and autonomous servo. The most important function, however, is to extract the required information from sensors for different applications such as physiotherapeutic rehabilitation, face recognition, gesture analysis etc.

The three dimensional (3D) information from a real scene of target objects can be compared with a pattern that may function as a basis for decisions. The pattern may also be renewed by the inclusion of a learning phase. These features require that the system can dynamically adjust the stereo camera to acquire the optimal 3D information. The IMSS can also contain multiple Radio-Frequency Identification, RFID, readers and tags which can easily be carried by a human being. Furthermore, the tags can be equipped with emergency buttons or different sensors that detect phenomena such as motion,

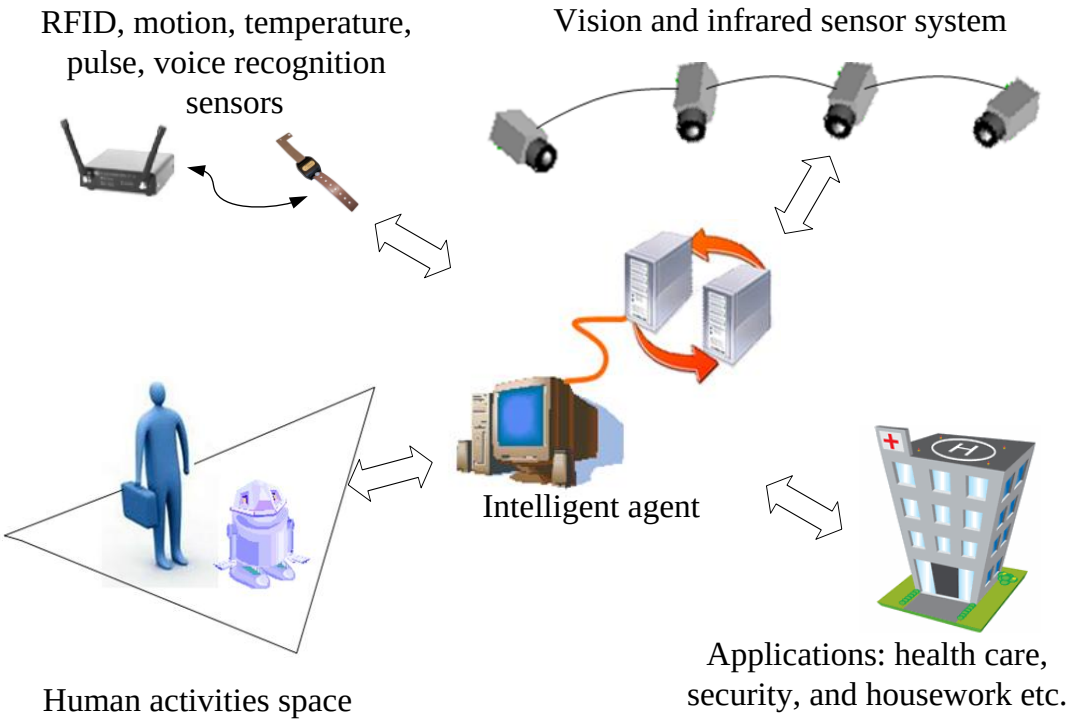


Figure 1.1. An overview of an Intelligent Multi-Sensor System

temperature, heart rate, as well as recognize voices. The RFID reader receives data from the tags that are used for person identification and initial localization.

The intelligent agent consists of a knowledge database that includes learning and decision making components that can be used to track, recognise and analyse the objects. The intelligent agent, functioning as the decision making unit, connects the different functionalities of the IMSS. As shown in Figure 1.2, its functionalities can be divided into six sections which are: collection and fusion, identification and positioning, extraction, enhancement, recognition, and finally decision making and control. These functionalities can be further described as:

- *Collection and Fusion*: information from different sensors has to be collected, fused and classified to different applications.
- *Identification and Positioning*: the target is identified, positioned and tracked using RFID, motion, and/or voice recognition sensors.
- *Extraction*: the information of the scene, including position and medical profile, is captured in order to define the target's characteristics.
- *Enhancement*: enhancing adaptive measurement methods using the integration of extracted qualitative factors and quantitative features are applied.
- *Recognition*: comparing the collected target's features with the pattern to recognise the target's state.
- *Decision making and control*: a decision about what action to take is based on pattern recognition. The decision can lead to different control steps.

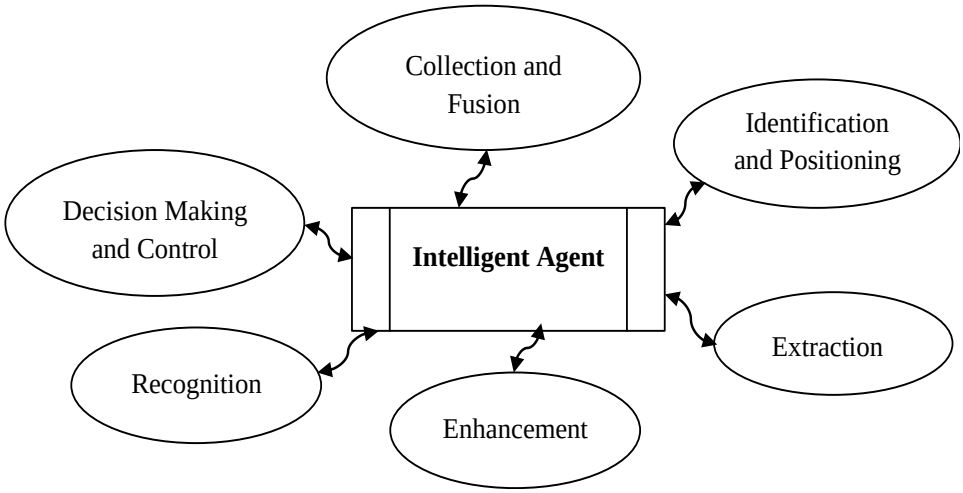


Figure 1.2. An overview of the relationship between the intelligent agent and the different functionalities as described in the thesis.

During the capturing of an image of a scene by a camera, explicit depth information about the scene is usually lost. However, depth information retrieval is one of the critical tasks for many applications in the IMSS, e.g. human motion analysis by 3D reconstruction can support monitoring and subsequent correction of human motions when patients perform rehabilitative training; robots can be navigated to assist people with housework and healthcare; human gestures and facial expressions can be studied for human-system interaction etc. Tracking the human position continuously and covering it by the camera's Field of View, FoV, is an essential problem in this system. The visual sensor configuration needs to be optimized to get necessary visibility and accurate 3D information.

Two images taken from the two cameras are obtained and then the 3D coordinates of an imaged point in the scene can be found by matching or computing the displacement, or disparity, between two corresponding feature points in the left and right images. There are two possible configurations of the stereo camera: parallel and convergent. For a parallel stereo camera, the corresponding points from the left and right images lie on the same horizontal scan-line, and then no rectification procedure is required. However, such a camera has a limited common FoV. By comparison, the convergent stereo cameras have a larger common FoV. However, in this case, matching requires a two-dimensional process which can be simplified by the rectification process. The rectification procedure transforms the corresponding points to be on the same horizontal scan-line. This means that the stereo matching algorithm reduces the search space from two dimensions to one dimension along the horizontal rows of the rectified images.

Accuracy is one of the major issues in 3D reconstruction and depth measurement, and this problem is addressed in Papers I – III. Due to the digital camera principle, the depth reconstruction accuracy is limited by the sensor pixel resolution which causes quantisation of the reconstructed 3D space. To improve accuracy, the quantitative analysis method of spatial quantisation using an iso-disparity map can be useful and this is shown in Paper I. The proposed mathematical model of the iso-disparity map provides an efficient way to describe the shape of the iso-disparity planes and evaluate a depth reconstruction uncertainty dependent on the stereo pair baseline length, the target

distance to the baseline, the focal length, the convergence angle, and the pixel resolution.

There are some methods which help overcoming limitations of the spatial discretisation. For instance, Kil et al., [3], have used a laser scanner to reconstruct a high-resolution 3D image of the target surface using hundreds of lower resolution scans as inputs. Another method is dithering which is a well-known technique applied in Analogue to Digital Converters, *ADCs*. This method decreases the system resolution below the least significant bit, [4], [5], [6]. By the guidance of an iso-disparity map, the introduced spatial dithering signal reduces the depth spatial quantisation uncertainty by half by combining four pairs of stereo images as shown in Paper II.

The sensor-shifted parallel stereo camera presented in Paper III was developed to provide an effect similar to the convergent stereo camera. The sensor-shifted camera yields a large common camera FoV, but still allows the corresponding points from two images to be easily identified without the stereo rectification procedure. The dithering method is also implemented and integrated with the sensor-shifted stereo camera to enhance the depth reconstruction uncertainty.

The target space, specified as a human activities space, requires the planning of the stereo sensors' configuration in order to increase the sensor observability for the optimal number of sensors. The planning of the sensor configuration by the use of reannealing software was introduced by Mittal, [7], and the evaluation of the sensors' configurations by a quality metric was presented in [8]. Furthermore, a linear programming method used to optimise sensor placement based on binary optimisation techniques has also been developed as shown in [9], [10], and [11]. This is a convenient tool to optimise the visual sensors' configurations when observing a target space such as a human activities space. Papers V and VI describe the optimisation programme for the planning of the stereo sensors' configuration used in the 3D reconstruction of a human activities space. The papers introduce a method by which the stereo pairs' configurations can be optimised under the required constraints of the stereo pair's baseline length, visibility, camera movement, and depth reconstruction accuracy.

During the tracking and surveillance process, the visibility of the targets is affected not only by the camera's FoV, but also by light, obstacles etc. To overcome this problem, radio technology can be useful. RFID is a rapidly developing technology based on wireless communication. The technology has been widely studied and is used in different applications, [12], [13]. The advantage of the RFID tracking system is its low-cost, large coverage area, independence of light, and ability to penetrate obstacles. Figure 1.3 shows an example where vision and RFID systems have been integrated in a home environment. The vision sensors are mounted on a mobile rig which can move along a track. The track is mounted around the upper part of the walls of the room and RFID sensors can also be arranged on it. The target motion can still be tracked even if the vision system is on standby or obstacles obscure the targets. The tracking information can be used for arrangement of the stereo vision sensor for an accurate measure of the target position. To estimate the position of the human within the studied space, Paper VI proposes the use of the Levenberg-Marquardt, LM, algorithm [14] with a Gaussian Mixture Probability Hypothesis Density, GM-PHD, filter [15] based on an RFID system by means of the tag's Radio Signal Strength, RSS.

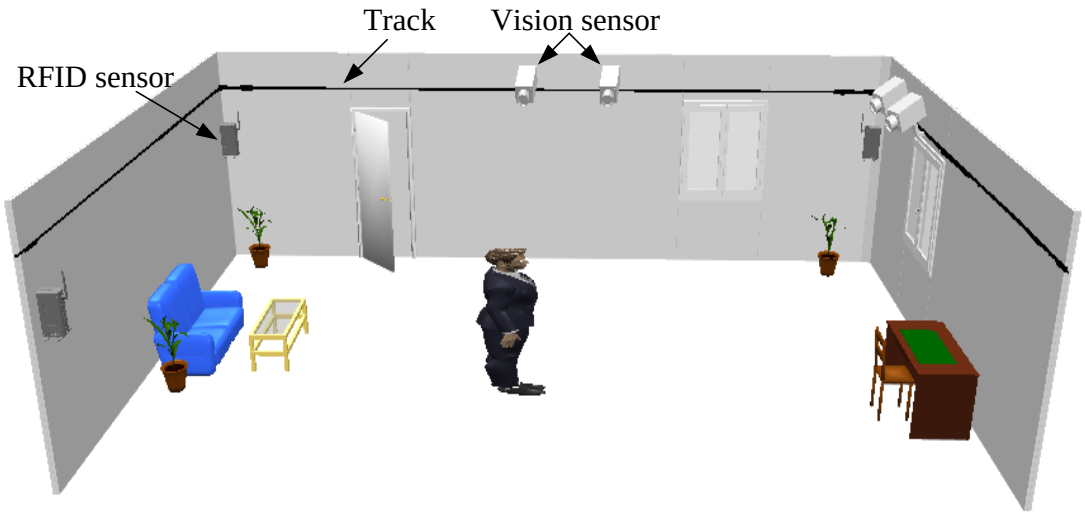


Figure 1.3. Multi-sensor for a human activities space integrated in a home environment.

Simulations and/or physical experiments for validation are used in all included papers. Two novel validation and evaluation methods are introduced. Validation of the reconstruction method aimed to improve the depth estimation accuracy is very difficult, mostly due to the problem of absolute depth measurement. However, a suitable and simple validation method that measures the differential depths between two points is proposed and implemented in Papers II and III. Paper VII proposes a method to evaluate the performance of the GM-PHD filter when tracking multiple targets. Motion speed and angular velocity of a circular test motion are the metrics suitable for the evaluation of the accuracy and label continuity of the tracking filter.

The evaluation of scientific and industrial measurements cannot be completed when there is a lack of traceable calibration of the specific type of measurements or instruments being used. The measurement of how the human being perceives image quality lacks traceable calibration. Therefore, this type of measured parameter cannot be compared with other measurements made using other methods, [16]. To address this problem, a quality index has been developed and used in different branches such as the food industry, [17], ecology, [18], and image processing, [19], [20], and [21]. The universal image quality index has been proposed by Wang et al., [20], [21], where the image quality index correlates with human visual perception. The proposed adaptive measurement method makes use of the Fuzzily Defined Variable, FDV, to reliably validate system performance. The FDV enables the combination of qualitative and quantitative factors into the evaluation procedure, [22]. The neural network as a tool is applied and uses learning and prediction functions to integrate the human factors into the IMSS. Image quality as it relates to human visual perception can be introduced as a modelling problem of the FDV, as shown in Paper VIII.

1.2 Thesis objective and scope

The objective of the research that this thesis accounts for is to develop, validate and evaluate the models and methods used in the IMSS introduced in the previous section. The implementations of these models and methods enable the observation and interpretation of the environment and the changes caused by human activities. To

achieve this goal, the research project has focused on the modelling of a human activities space. In connection with this, technologies and methods for human tracking, the control of vision systems, depth reconstruction improvement methods and finally validation and performance evaluation methods used for depth measurement and human tracking systems have been employed.

The scope of the thesis can be described as follows:

1. The mathematical model for the general stereo pairs is proposed and analysed. The depth reconstruction uncertainty is represented in the form of the distances between the iso-disparity surfaces. The depth reconstruction uncertainty for parallel stereo cameras is analysed. A sensor-shifted parallel stereo camera is used to enlarge the common camera FoV. The dithering algorithm is introduced and applied to reduce the depth reconstruction uncertainty.
2. For tracking and surveillance purposes, the human is modelled as a tetrahedron. The sensors' intrinsic and extrinsic characteristics are considered while optimising the positioning and configuration control of a multi stereo visual sensor that assures accurate 3D reconstruction of the real scene. The RFID system is used for initial localisation and for tracking in case the vision system cannot be applied.
3. The validation method of the 3D reconstruction is developed and implemented. The performance of the tracking system is evaluated by applying the circular test motion. The image quality assessment carried out by means of the integration of quantitative and qualitative factors, is implemented by a neural network.

1.3 Thesis outline

The work presented in this thesis is based on the eight papers reproduced in Part II. These papers contribute to the three aspects of the IMSS: depth measurement, human tracking and surveillance, and the validation and evaluation of the proposed models and algorithms. Figure 1.4 illustrates the placement of the papers in respect to these three important aspects. In the figure, three aspects formulate three coordinates which are: depth measurement, tracking and surveillance and validation and evaluation, respectively. Each paper, represented by a cube, is placed in the coordinate space according to the percentage of the contents related to the three aspects.

Papers I to III are mostly related to the depth measurement aspect, a concern estimated to make up more than 50% of the paper. For the case of a pair of stereo cameras, Paper I focuses on the depth reconstruction method and the estimation of the depth measurement uncertainty. It is shown how the baseline length, sensor resolution, convergence angle, and the distance between the target and the camera influence the depth measurement uncertainty. For a convergent stereo camera, the 3D reconstruction uncertainty is presented with the aid of an iso-disparity map. The improvement of the 3D reconstruction accuracy by use of the dithering algorithm is introduced in Paper II. The sensor-shifted stereo camera, which combines the advantages of the parallel and the convergent stereo cameras by simplifying depth reconstruction without compromising the FoV and the accuracy, is proposed in Paper III.

Papers IV to VII focus on the human tracking and surveillance aspect, which is estimated to make up more than 40% of the contents. Paper IV proposes to track a

human position using the LM together with the GM-PHD algorithms in a RFID system. Paper VII also presents human tracking by means of the GM-PHD algorithm, but in this case, a stereo vision system is applied. In papers V and VI, the human is modelled by a tetrahedron to facilitate a control algorithm optimising continuous target tracking and 3D reconstruction. The optimization concerns the number of stereo cameras, its baseline lengths and its positions and orientations.

The common subject of all papers is the validation and evaluation aspect. 50% of the content of Papers VII and VIII are related to this aspect. Paper VII proposes an evaluation method of the tracking performance. Paper VIII deals with the validation methods used in quality assessment through integration of qualitative factors and quantitative features. 30% of Papers I - III and 20% of Papers IV-VI are concerned with this aspect.

The thesis consists of two parts, where Part I provides a general overview of the subject and methods of the thesis, and Part II presents the published papers.

The aim of this first chapter of Part I is to provide a brief overview of the relevant research areas and methods used in the thesis. Chapter 2 describes the research methodology behind the project. In Chapter 3, the measurement methods used in the project as well as their uncertainty and reliability are analyzed. Chapter 4 describes human tracking in the RFID and vision systems respectively. The planning of multi stereo sensors configurations used to monitor a human activities space is presented. Chapter 5 focuses on validation and evaluation methods. A brief summary of the included papers, the conclusion of the thesis, and suggestions for future work are given in Chapter 6.

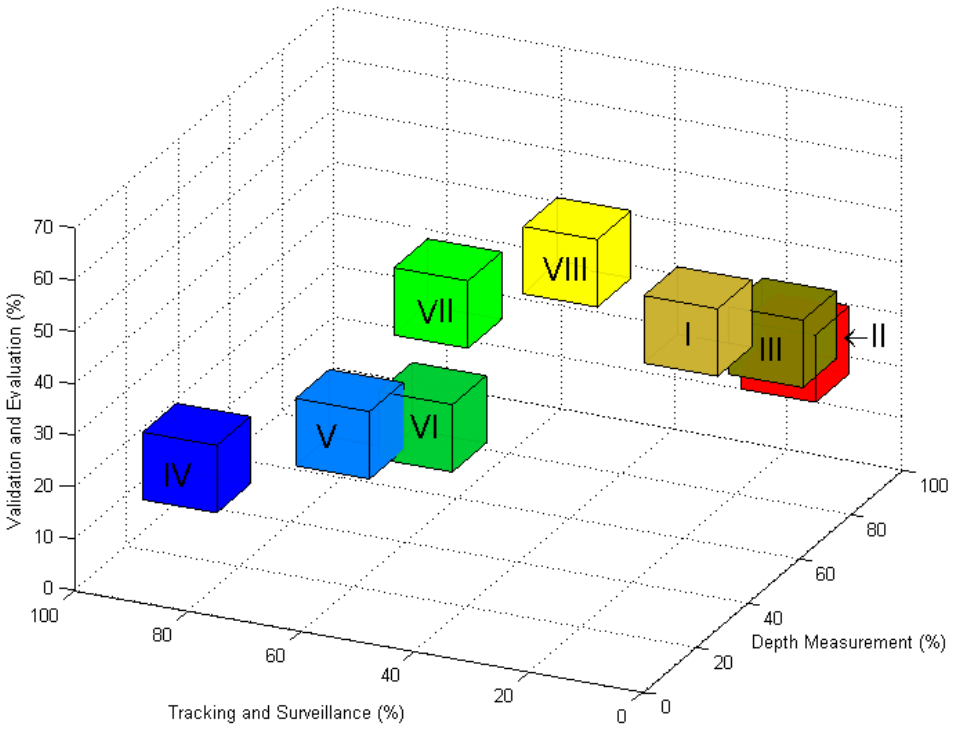


Figure 1.4. An overview of the papers presented in the thesis, related to three content aspects: depth measurement, tracking and surveillance, and validation and evaluation.

2 Research methods

This thesis concerns theoretical and applied research related to an intelligent multi-sensor system used to monitor a human activities space. The system performs a diversity of sensing functions including the acquisition, capture, communication and analysis of information. The information acquired by the system concerns the state of the human appearance, the health situation, activities, the surrounding environment etc. The information may be captured and communicated in a variety of signal forms. Since the human factor is an important part of the system, the integration of quantitative and qualitative paradigms is required, [23]. However, the main part of the thesis is based on quantitative engineering research methods, consisting of several phases, [24]: problem identification and inquiry, hypothesizing a solution; solution method development and implementation; and finally validation and evaluation of the solution.

2.1 Problem identification and consideration, hypothesizing a solution

The research work starts from the formulation of a question that *identifies the problem*. After considering the problem, the problem or research question is reworded into a hypothesis that can be tested by an experiment. "A good hypothesis states as clearly and concisely as possible the expected relationship (or difference) between two variables and defines those variables in operational, measurable terms", [25].

Finding the solution moves the problem from a given state to a new desired state. According to known problem-solving techniques, solution finding can be summarized by the following steps, [26]-[29]:

1. Division of a complex problem into sub-problems with isolated related factors;
2. Review of previous discoveries already made in the same or related fields;
3. Finding a hypothesis that may constitute a solution to the research question or problem.

All presented papers make use of those problem definitions and formulations to address the research questions that we then move on to solve. Papers I - III focus on the problem of depth reconstruction and its uncertainty in general. Paper I answers the question of what the relationship between the 3D reconstruction uncertainty and the pixel size, the focal lengths, baseline length and convergence angle is. Papers II and III present solutions on how to overcome the sensor resolution limitation and enable more accurate depth reconstruction based on the dithering algorithm. Papers IV - VI focus on visibility, accuracy and continuity during tracking and surveillance of a human activities space. Paper IV shows how to track a target even when it is obscured by obstacles. Papers V and VI address the problem of how to optimise the number of cameras and

their corresponding positions and orientations when observing the human body and its activities space. Paper VII evaluates how the target speed motion and angular velocity affect the GM-PHD filter's tracking performance in a stereo vision system. Paper VIII discusses how the adaptive quality model handles the quantitative and qualitative factors to assess quality measurement when there is a lack of traceable calibration in the measurements or instruments.

2.2 Solution development and implementation

New theories, models, algorithms and tools need to be developed to solve complex engineering problems. The approach we suggest requires a quantitative and qualitative analysis of engineering problems. In general, possible solutions can be categorized as:

- Developing new theories, algorithms and models;
- Applying existing algorithms to a new area;
- Combining existing methods and techniques in a unique way.

We combine different solution categories to solve the problems. For example, Paper I applies the iso-disparity model for analysis of the depth measurement uncertainty. This model is also applied in Papers II and III where the well-known dithering algorithm is applied to enhance depth reconstruction and measurement. Meanwhile, paper III combines a sensor-shifted camera and the dithering algorithm to simplify 3D reconstruction complexity without compromising the FoV and the depth reconstruction accuracy. In Paper IV, we combine the LM algorithm with the GM-PHD filter in a unique way for multi-target tracking using an RFID system. Papers V and VI propose a novel tetrahedron model for modelling the human in the activities space. A new approach to model the camera's FoV in spherical coordinates is also implemented. Moreover, the papers creatively apply the greedy algorithm with different stereo constrains to optimise the visibility of the stereo vision system in order to ensure accurate depth reconstruction. In Paper VII, the new evaluation method that makes use of a standardised circular test motion for a multi-target tracking system is proposed. Paper VIII implements the neural network to merge qualitative and quantitative factors, and thus solves the problem of combining such factors in a new way.

An engineering solution requires development of a practical method that can implement theories, models and algorithms into a real system. The most common issues that must be addressed when implementing a theory, a model and an algorithm are reliability, efficiency, and complexity. *Reliability* implies that the method and its implementations are robust and stable. *Efficiency* means that the method is realizable with a high-performance in regards to execution time, used memory and energy. *Complexity* stands for implementation simplicity and clarity.

In Paper II and III, where the dithering algorithm is applied to enhance the depth reconstruction uncertainty, the implementation consists of four steps. Firstly, the depth of the target point is preliminary roughly measured and secondly the dithering signal is estimated. Thirdly, the depth of target point is estimated again by a new disparity after applying the dithering signal. Finally, the calculation of the depth of the target point is done.

When implementing tracking and surveillance, the simplicity, efficiency and reliability of the integer linear optimization were the implementation issues that affected the cameras' constraints. The binary variable used to manage visibility is computed and stored in advance to reduce the computational burden during the real-time process. The binary variable also makes calculation simple and stable. In Paper VII, the implementation of a multi-target tracking algorithm requires that an affixed label is assigned to each target. If a new target appears, a new label is added to the set, and similarly if a target disappears, its corresponding label is discarded from the label set. The filter can detect and handle the targets' label discontinuity to keep tracking reliable.

2.3 Validation and evaluation method

Kuhn has claimed that the measurement plays an important role in quantitative research, [30]. The observations are expressed as numerics in the measurement procedure. If the purpose of the measurement is to acquire data about the dynamic behaviour of the real targets, it can be referred to as *inferential measurement*. Such a measurement is used to determine the dynamic behaviours of a process such as the ability of a system to process, store, transform, and transmit data, [31]. Validation and evaluation methods are the inferential measurement used to verify correctness and evaluate the performance of the analytical model and the procedure employed for a specific task. Validation and evaluation results allow an estimation of the quality, reliability, efficiency, complexity and consistency of the analytical results. The measures used for validation and evaluation are based on simulation and/or real experimental tests, [16].

The measurement uncertainty must be considered when validation and evaluation are concerned. The *measurement uncertainty* indicates the range of values within which the true value is estimated to exist. The uncertainty can arise from different sources, and can be described in the following way, [16], [31]:

- The method uncertainty may be caused by standard, calibration or instrument characteristics;
- The procedure uncertainty may be caused by the characteristics of the measurand and the interface between the human being and the instrument;
- The measurement environment uncertainty may be caused by changes in temperature, pressure, humidity, the power supply etc.

In Papers II and III, both simulation and real experiments are used for method validation and the real experiment results have a larger uncertainty range than the simulation due to additional factors such as camera calibration, lens distortion etc. The depth measurement differential method is applied to validate the depth reconstruction model. In Paper VII, the 3D circular motion test signal is introduced to evaluate the tracking performance according to the target speed and the angular velocity. The test signal with its standardised characteristics makes it possible to compare different methods and their efficiency for multi-target tracking. When considering the human factor in the system, both the human being and the technical aspect must be considered when evaluating the quality of the system. Paper VIII introduces qualitative factors into the evaluation process and quality is described with the help of a Quality Index, QI, which rely on different kinds of quantitative and qualitative parameters.

3 Measurement - methods, uncertainty and reliability

3.1 Depth measurement and 3D reconstruction - spatial modelling and uncertainty analysis

How the human being is able to view the world in three dimensions has been a concern for a very long time. During the seventeen century, the question was routinely phrased as: *How does human depth perception work?* The answer to this question led to stereoscopy. Nowadays, a stereo camera is routinely used for 3D imaging. The 3D reconstruction of a scene from images has been studied for many years in photogrammetry and computer vision. There are many different methods which have been developed and used in the 3D reconstruction of buildings, human faces, industry products, etc. Finding the depth of a point in the scene is the most important task in 3D reconstruction. In order to determine the 3D position of a specific point, one needs at least two images. The necessary information regarding depth and the relations between objects can be found using those two images.

Figure 3.1 shows the principle of using two images, taken by a convergent stereo

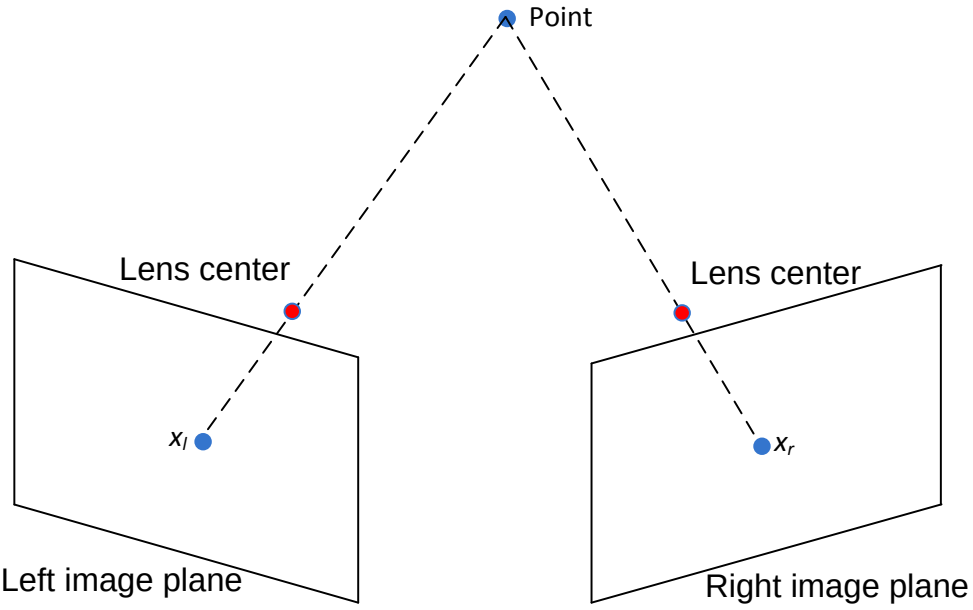


Figure 3.1. The point in 3D space can be reconstructed by a triangulation method.

camera and used to reconstruct a point in a 3D space through a triangulation method. If we can observe the same points from two different positions, we can deduce two rays from the left and the right camera centre and their corresponding projection points in the images. The intersection of the rays defines the point location in space. The 3D reconstruction from the two views is based on an epipolar geometry which describes the relationship between the corresponding image and the scene points. In order to obtain the 3D information, the image points' coordinates and the camera configurations are needed. The 3D reconstruction procedure essentially consists of the three following steps, [32]:

- Matching: finding the corresponding image points for the same scene point;
- Calibration: obtaining the position and orientation of the stereo camera for the different views;
- Reconstruction: extracting the relation between the image points and their corresponding rays.

During the last reconstruction step, the relation between the image points and their corresponding rays is obtained from the pinhole camera model. This model is defined by the intrinsic and extrinsic parameters of the camera, [33]. The *disparity*, the quantity used in depth reconstruction, refers to the displacement of the corresponding points on the left and right images for a common scene point along the corresponding epipolar lines.

The iso-disparity surfaces characterise the quantisation phenomena in stereo reconstruction, [34], [35]. The intervals between the discrete iso-disparity surfaces represent the depth reconstruction uncertainty. The iso-disparity surfaces' geometry models proposed in Paper I are valid for both common configurations of a camera

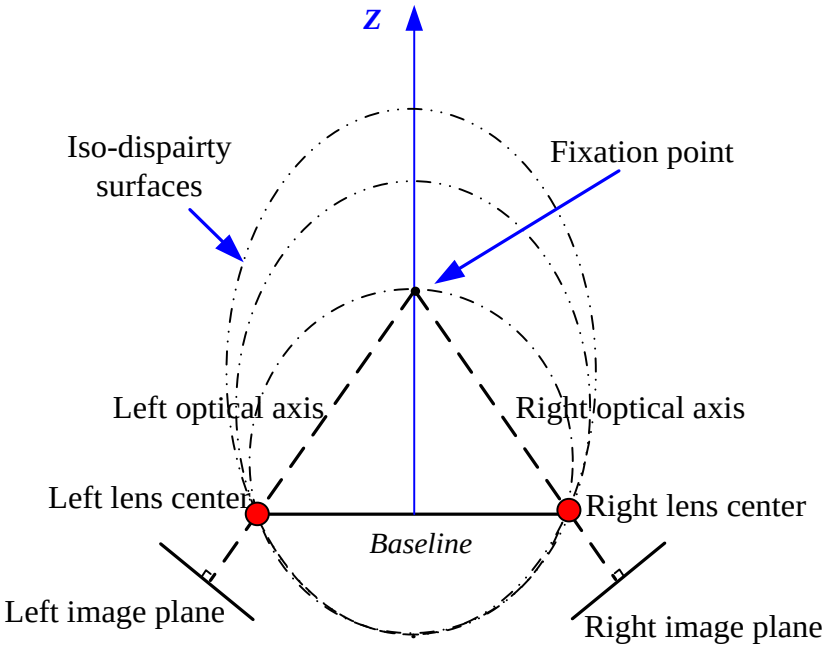


Figure 3.2. An illustration of the iso-disparity surfaces' geometry model for the convergent stereo pair in the plane defined by the cameras' optical axes.

stereo pair: the convergent stereo pair and the parallel stereo pair.

In Paper I, the mathematical model of the iso-disparity surfaces has been analysed for a convergent stereo pair for the most general common configuration, where the optical axes cross at a fixation point, as shown in Figure 3.2. The zero disparity circle is defined by the fixation point and the left and right cameras' optical centre position points. This circle is known as the Vieth-Müller circle and is a projection of the horopter, [36]. The iso-disparity surface of the quantised disparity for a convergent stereo pair with the same focal length and the same convergence angles describes a cylinder, where the ellipses are cross sections of this cylinder on the optical axis plane. In order to define the ellipse's position, shape and orientation, we need to define the ellipse's five degrees of freedom. This is described in Paper I, which presents a convenient way to analyse the depth reconstruction accuracy.

The second common configuration is a parallel stereo pair in which the optical axes of the cameras are parallel. This could be considered as a special case of the convergent stereo pair's configuration with the fixation point set to infinity. The cameras may have the same focal lengths, or their focal lengths may be different, e.g., to get a better reconstruction accuracy of a target placed at the boundary of the cameras' FoV, [35]. The geometry model shows that the iso-disparity planes are parallel for a parallel stereo pair with the same focal length, while the iso-disparity planes intersect at a straight line for the parallel stereo pair with different focal lengths. The plots of the iso-disparity planes for these two configurations of the parallel stereo pair are shown in Paper I.

The depth uncertainty analysis of the target space and the corresponding algorithm for optimising the number of stereo pairs and the stereo camera's configurations are presented in Paper II. The depth reconstruction accuracy depends on the system configuration which is defined by sensor resolution (pixel size), focal length, baseline length, and convergence angle. The depth reconstruction uncertainty is described by the iso-disparity geometry model and varies significantly with respect to the target distance to the baseline, the baseline length, and the focal length. However, when determining the accuracy of a 3D reconstruction, the depth spatial quantisation caused by a discrete sensor is one of the most influential factors. This type of uncertainty usually cannot be decreased by reducing the pixel size because of the restricted sensitivity of the sensor itself and the declining signal to noise ratio this would lead to.

By adjusting the stereo pair's profile, such as the baseline, the focal length, and the pixel size, the depth reconstruction accuracy can be improved. The depth spatial quantisation factor is one of the most influential factors when determining the accuracy of a 3D reconstruction.

3.2 Improvement of the depth measurement

To show how the depth measurement and 3D reconstruction accuracy can be improved, the iso-disparity surfaces' geometry model and the dithering algorithm are presented in this chapter. First, the depth reconstruction uncertainty in relation to different setups, e.g., parallel and sensor-shifted stereo cameras, is discussed and the corresponding algorithm for enhancing the depth reconstruction uncertainty is applied.

How to reconstruct a super-resolution image from the low-resolution images has been the focus of much research in recent years. To overcome the digital camera sensor

pixel size limitation, attempts have been made to combine the information from a set of slightly different low-resolution images of the same scene and use them to create a higher-resolution image, [37], [38]. The selection of an optimal sensor pixel size is discussed by Chen et al., [39].

When compared to the convergent stereo camera setup, the sensor-shifted stereo camera setup makes use of a simpler reconstruction process that does not require rectification processing. In addition to this, the sensor-shifted camera offers a wider common FoV than the parallel camera stereo setup. Fig. 3.3 shows the common FoV as a blue shaded area for the parallel stereo pair and as a pink shaded area for the sensor-shifted parallel stereo pair respectively. This shows that the sensor-shifted parallel setup has a wider common FoV than the parallel stereo setup. Francisco and Bergholm suggested the use of a sensor-shifted camera in the stereo setup, where the sensor has a controlled micro-movement, [40].

Signal processing methods can improve accuracy. Dithering is one such method, and the usefulness of this method is explored in Paper II and III. In our proposed model, the left and right cameras are the quantisers. The dither signals add *noise* to the signals (the projections of a scene point) prior to its quantisation in order to change the statistical properties of the quantisation, [5]. In our case, there are two possibilities to add a dither signal to change the projection positions: one is to shift the target features parallel to the image planes and an alternative is to shift the camera sensor, which means that the quantisation levels of the quantiser are changed.

The dither signal is discrete and is used to control the left and right cameras' positions. We have presented a two-stage discrete dither signal for each camera, which provides four images from which to calculate the depth of the target feature with an improved resolution and a reduced quantisation uncertainty. From Paper II, we know that the optimal dither signal makes the target projection move from its original position

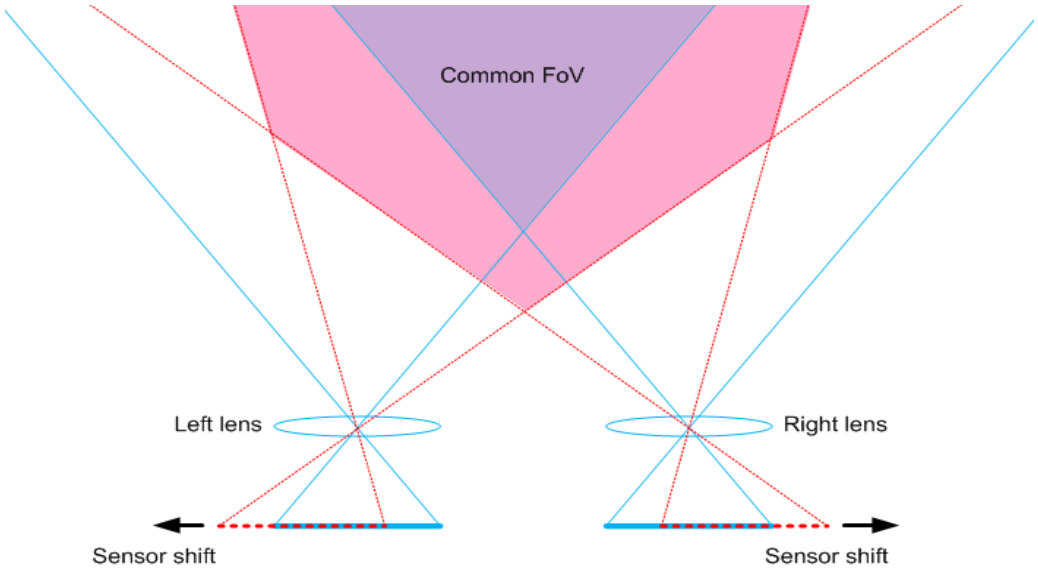


Figure 3.3. The common FoV where the blue solid lines represent the parallel stereo setup and the red dashed lines represent the parallel sensor-shifted stereo setup. Shaded areas are the common FoV for the parallel and sensor-shifted stereo setups respectively.

by a distance that is equal to half a pixel size.

The iso-disparity surfaces become parallel planes for the parallel stereo cameras, having a convergence angle equal to zero. The depth reconstruction uncertainty can be reduced by half when the dithering signal changes the camera position. Actually, the phenomena can be explained as the dithering signal moving the new iso-disparity planes into the middle between the primary disparity planes, as described by Figure 3.4. As we can see, by combining secondary and primary iso-disparity planes, the uncertainty is reduced by half. Iso-disparity planes can be moved by increasing or decreasing the baseline or by shifting the sensor chip. In Paper II, the baseline change is applied and this is accomplished by a single camera movement. To change the baseline length by placing the new iso-disparity plane in the middle between the primary iso-disparity planes is also the optimal solution from a quantisation point of view. Paper II provides an example of a change in the baseline length and of how this change can be analysed. The movement of the iso-disparity planes is effected simply by shifting the sensor chip in Paper III. The shifting of the sensor chip by half a pixel size by the dithering signal is in effect to place the new iso-disparity plane in the middle, between the primary iso-disparity planes. The implementation of the dithering algorithm on the sensor-shifted stereo camera is also shown in Paper III. The simulation and the results of the real experiment in Paper II and III confirm the analysis. The depth reconstruction uncertainty is reduced by half.

3.3 Adaptive method for the measurement of fuzzy quantities

The adaptive method is needed to measure a quantity which is significantly affected by the human factor. Human characteristics vary from person to person and depend on many factors. Such quantities usually cannot be precisely defined and do not possess any physical standards. To model and measure them, one can fruitfully use methods which are able to adapt themselves to individual personal characteristics. Such

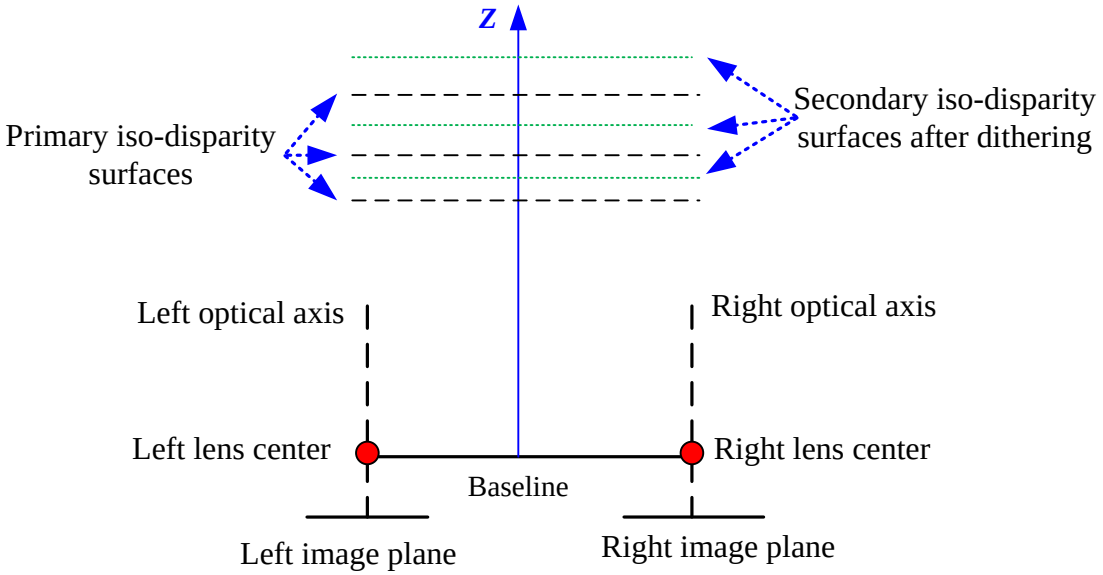


Figure 3.4. The combination of primary and secondary iso-disparity surfaces for the parallel stereo pair in the plane defined by the cameras' optical axes.

quantities are defined as *fuzzy*, and they are discussed in Paper VIII with the aid of the Fuzzily Defined Variable, FDV.

Since the FDV often depends on both quantitative and qualitative factors, it is difficult to express it in only quantitative terms, [41], [42]. The FDV factors are of different importance for different targets or users. The attributes of the FDV are not clearly defined, since they depend on different types of features and factors. The choice of suitable features and factors depends on the target group, and/or the cultural environment, and/or the age, and/or education, etc., within the application field. Because of this, the two main dependencies that must be handled within the FDV are related to:

1. The set of features that are a part of the FDV and depend on:
   a. Expertise;
   b. Possible measurements;
   c. Pattern data.
2. The weights of the FDV that depend on:
   a. Human perception - assessment;
   b. The feature's relevance;
   c. *Measurement uncertainty;*
   d. *Other factors such as cost or complexity.*

As a way to measure the FDV, we propose a quality index that is created through an adaptive method. The quality index can be generalised for many different kinds of purposes and the measurement method can be adjusted to changeable parameters.

3.4 Measurement and prediction uncertainty and reliability

As considered in this thesis, the measurement process takes the input quantity and transforms it into an output quantity that can be observed or used. The accuracy of a measurement process can be described as the degree of closeness of the measured value to its actual value. The accuracy is associated with measurement methods, tools, environments and human and other factors. The measurement data analysis can be qualitative or quantitative. The quantitative analysis applies mathematical models, both deterministic and statistical, and is implemented in Papers II and III. The prediction uncertainty acquired by means of the Wasserstein distance is shown in Paper VII. The qualitative analysis uses statistical or fuzzy models and is applied in Paper VIII.

Since the camera sensors can be considered as mid-tread quantisers, the projection of the scene point is approximated by the pixel center. The difference between the approximate and the exact projections is the image quantisation uncertainty. In our approach, we are interested only in the image's horizontal uncertainty. The probability distribution functions of the images' horizontal quantisation uncertainties for the left and right images are rectangular. The disparity quantisation uncertainty, as the result of the convolution of two rectangular distribution functions, is triangular. The depth reconstruction quantisation uncertainty is the non-linear function of disparity and corresponds to the interval between the iso-disparity surfaces. An example of the depth quantisation reconstruction uncertainty for the parallel stereo camera is shown in Figure 3.5.

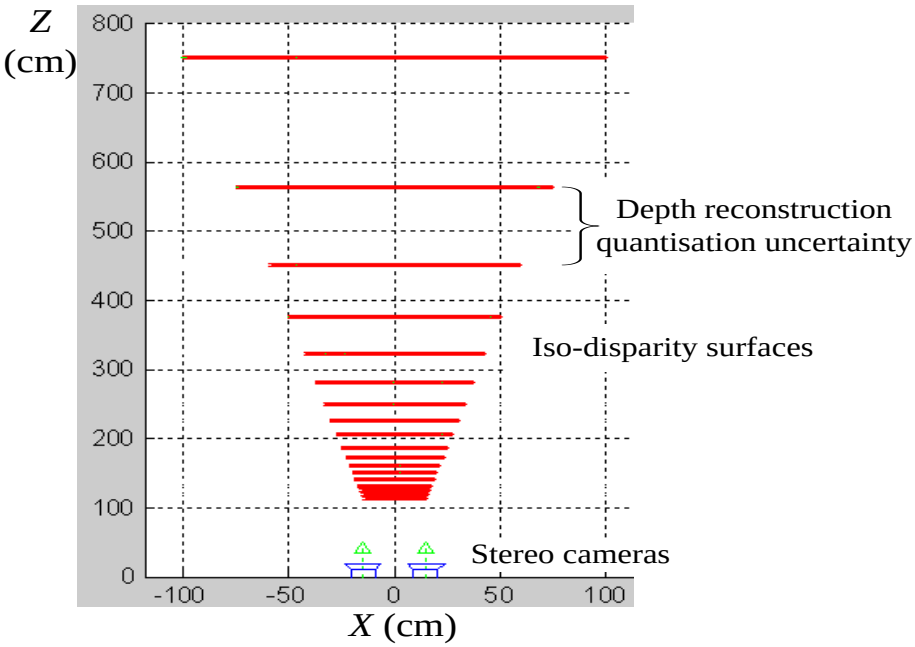


Figure 3.5. The depth reconstruction quantisation uncertainty is represented as the interval between the iso-disparity surfaces.

The analysis of the improvement of the spatial depth quantisation when using the dither signal as presented in Papers II and III is an example of the statistical model in quantitative analysis. In our approach, the Probability Density Function, PDF, of the discrete dither signal requires equiprobable impulses. In our case, there are two impulses, so a two-stage discrete binary dither signal can be applied. The statistical analysis presented in Paper II shows that the mean value of the image's horizontal uncertainty has a minimum interval when the dither signal is equal to half a pixel.

The prediction uncertainty for multi-target tracking by the GM-PHD filter is measured by the Wasserstein distance adopted from theoretical statistics as a means of defining a metric for multi-target distances, [43]. The mean value of the absolute error defined as the difference between the prediction and the ground truth is measured by using the Wasserstein distance in Paper VII. This enables us to evaluate the reliability of the filter.

The human factor has been considered in the systems where a qualitative analysis has been performed. In order to avoid the problem that quantitative methods tend to overwhelm qualitative ones, each component needs to be treated as parts of a puzzle, [16]. Measurement reliability may rely on both quantitative and qualitative factors. These can be presented as an Ishikawa diagram, [44], see Figure 3.6. Measurement reliability is influenced by quantitative factors such as target properties, measurement uncertainties, methods, and tools. However, the reliability also depends on many qualitative factors such as the psychical environment, target background and usefulness, all related to the target and the human being. A measurement of the image quality by the combination of quantitative and qualitative factors is presented in Paper VIII.

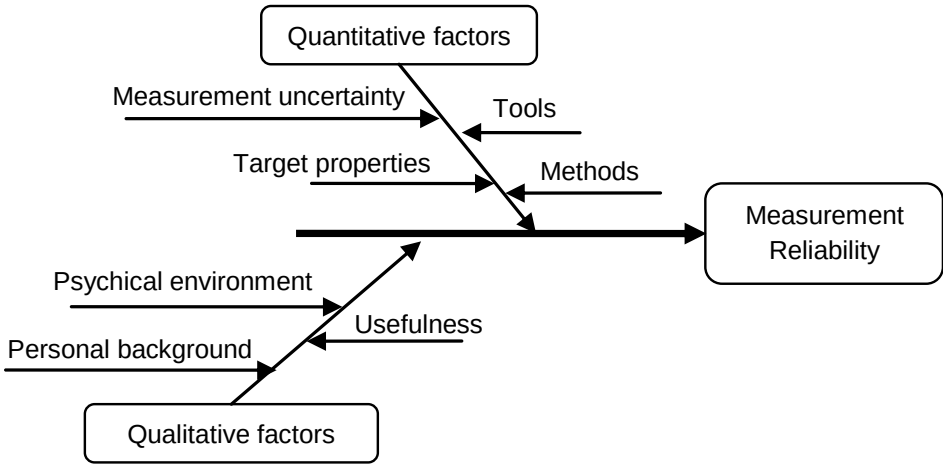


Figure 3.6. Measurement reliability may rely on quantitative and qualitative factors.

4 Tracking and surveillance of a human activities space

Similar to the human eyes, the multi-vision sensor system acquires information about a target's colour, shape, position etc. in a 3D space. A human activities space, as a target space for 3D reconstruction, requires tracking and monitoring. The motion and possible locations of the human being need to be predicted in order to control the monitoring system so that the target can be kept in the cameras' FoVs.

The constraints for the design and planning of the active stereo vision system can be determined from the human activities space and the vision sensor parameters which guarantee the optimal visibility and required 3D reconstruction accuracy in the vision sensor system. The sensor planning can be viewed as an extension of the well-known Art Gallery Problem, AGP, [45]. The AGP describes a simple polygon, often with holes, and the task is to calculate the minimum number of guards necessary to cover a defined polygon. When researching a human activities space, the goal can be re-defined as: find the minimum number of stereo pair sensors needed to cover the target space.

To optimize sensor placement, many researchers present the required coverage space as a regular grid, [9], [10]. The optimization algorithms proposed by them function in a 2D space, but the approach presented in this thesis applies the optimization algorithm to a 3D space, [46]. In the field of active vision, there have been some studies on how to apply the dynamical adjustment of the stereo pair baseline to improve the reconstruction accuracy, [47], [48]. However, as Mittal mentions, there has been relatively little work on determining the optimum sensor configurations which is what our research addresses, [7].

The Bayesian's recursion provides the mathematical framework for the tracking algorithms. A typical tracking algorithm, applying the Bayesian's recursion, computes the posterior probability density of a process based on the prior probability density of the process and the likelihood function. The GM-PHD, algorithm was proposed by Vo and Ma [15]. Using the GM-PHD algorithm, Pham et al. have proposed a multi-camera, multi-object tracking system that can track 3D object locations even when objects are occluded from the cameras, [49]. In [50] they have demonstrated how efficient it is to track the multi-target by using a multi-sensor.

The visibility of the targets is affected by light, obstacles, and the camera's FoV etc. The human factor or cultural considerations can also interfere with the vision system. For instance, there may be privacy protection in a certain area. Complementary technologies which are independent of such interferences can support surveillance and tracking. For example, the tracking of a specific target can be achieved by using positioning systems like the Global Positioning System, GPS, ultrasonic signals,

infrared and radio frequency signals such as RFID technology, [51], [52], [53]. The localization techniques used in RFID rely mostly on an accurate estimation of the distance between the reader and the tag. The advantage of the RFID tracking system is its low-cost, large coverage area, independence of light, and ability to penetrate obstacles. It is especially important for such cases when the vision system is on standby due to occlusions, darkness or due to personal reasons.

This chapter provides an overview of the tracking and surveillance methods that can be used in a multi-vision sensor system. The GM-PHD algorithm, able to predict the target position using RFID and vision systems, is presented in Paper IV and Paper VII, respectively. The prediction can be used for the camera arrangement algorithm used in tracking. Papers V and VI introduce the human model as a tetrahedron and the camera FoV model using spherical coordinates to get optimal visibility of the target. Furthermore, the iso-disparity geometry model allows an accurate analysis of the depth reconstruction process. A depth reconstruction accuracy constraint is proposed and applied in the model. The human and camera FoV models and accuracy constraint aid the process of dynamically adjusting the positions, orientations, and baseline lengths of multiple stereo pairs of cameras, thus acquiring the desired visibility and accuracy.

4.1 The surveillance method and its implementation

For surveillance, the constraints for the arrangement of the stereo camera should be considered in relation to the human activities space, the camera FoV, and the depth reconstruction accuracy requirement. The optimal number of cameras and their corresponding positions and orientations when observing the human body and activities space can be calculated using the integer linear programme. The constraints applied in the integer linear programme can be defined to obtain sufficient visibility and enough accuracy for surveillance, tracking and 3D reconstruction.

How the sensors should be placed for optimal visibility has been studied by many researchers, [7], [8]. However, their work focuses on the target area coverage problem without specifying a target within this area. In our work, the target is the human. We propose that the target object should be modelled as a tetrahedron. This method provides a convenient way to extract the orientations of the surfaces that represent the front and back of the human being. The tetrahedron model constraints guarantee that the human body can be viewed from each of these directions with the desired image resolution.

The human modelled as a tetrahedron is shown in Figure 4.1. The normal of each tetrahedron's upper triangle gives the orientation of that surface. If the visibility angle, θ, between the triangle normal and a line drawn from the centroid of the triangle to a specific camera position increases, then the image resolution decreases. In order to get good image resolution, a visibility angle, θ, of less than the maximum visibility angle, θ_{max}, is required.

The camera orientation should line up with the centroid of the triangle, thus bringing the target object to the centre of the camera's FoV and causing less lens distortion. The angle between the camera orientation and the line drawn from the camera position to the centroid of triangle, φ, must be less than the maximum angle, φ_{max}.

Since stereo matching becomes more difficult when the baseline distance increases, the baseline length b has to be limited to the maximum stereo baseline length, b_{max}.

Due to the limited dynamics of the camera, a camera movement distance constraint can be applied. The next-view position of the camera should not be placed too far away from the previous position. This constraint is formulated as the camera maximum movement distance, and it should be less than the maximum camera movement distance that the system supports:

$$Dist(StereoPair_{next}, StereoPair_{current}) \le Dis_{\max}, \tag{4.1}$$

Hörster and Lienhart use a simple model for their cameras: the FoV of a camera is described by a triangle and applied to a 2D room layout map, [10]. Our camera FoV model works in a 3D space and is thus able to cover a 3D human model. Modelling the camera's FoV using spherical coordinates simplifies the model and the constraints, thus speeding up computations.

Then, the spherical coordinates are applied; the camera horizontal and vertical viewable angles, ϕ_h, ϕ_v, and a working distance, r, can be calculated from the camera's attributes. The spherical coordinate system is shown in Figure 4.2. In order to keep the target object's feature points within the camera's FoV, the following constraints must be fulfilled:

$$\begin{gathered} l \le r \text{ and} \\ \alpha_c - \phi_h/2 \le \alpha_o \le \alpha_c + \phi_h/2, \\ \beta_c - \phi_v/2 \le \beta_o \le \beta_c + \phi_v/2, \end{gathered} \tag{4.2}$$

where l is the distance between the target position and the camera's position; α_o, β_o are

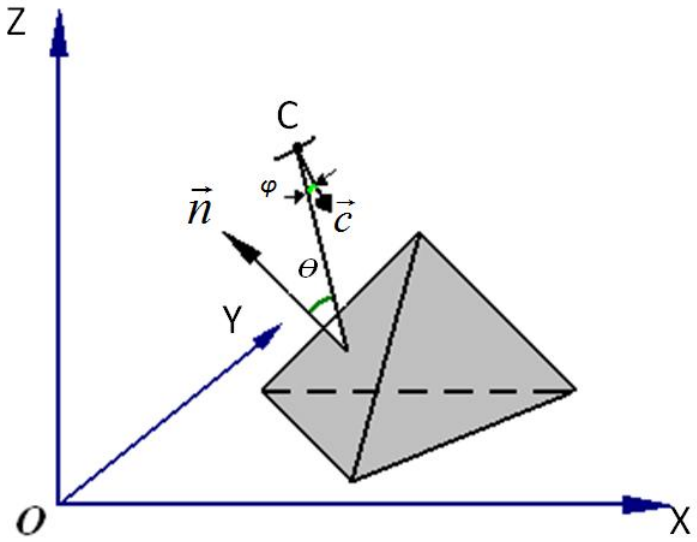


Figure 4.1. Illustration of the human modelled as a tetrahedron; θ is the visibility angle between triangle normal $\vec{c}$ and a line from the centroid of the triangle to the camera position; φ is the angle between the camera orientation $\vec{c}$ and a line from the camera position to the centroid of the triangle.

the azimuth and the elevation of the target, respectively; and α_c, β_c are the azimuth and the elevation of the camera's orientation, respectively.

Improvement of the depth reconstruction is one of the major focuses of this research project. This thesis suggests that a depth accuracy factor, *AF*, is a function of the target's convergence angle, ψ, and the camera's orientation, α_c. In fact, *AF* varies more significantly in respect to the target's convergence angle than in relation to the camera's orientation. Thus, the target convergence angle determines the depth accuracy factor. The accuracy constraint for a given point can be defined as:

$$AF \le AF_{con}, \tag{4.3}$$

where AF_{con} is determined from the reconstruction accuracy requirements of the given application.

The integer linear algorithm gives an optimal solution of the number of cameras needed to observe the target with the required constraints. The algorithm assumes that one type of camera is used throughout, so that all cameras' FoV ranges are the same. The integer linear algorithm can be applied to minimize the total number of stereo pairs subjected to the FoV coverage, the baseline length, the camera maximum movement distance and the depth reconstruction constraints. This objective function minimizes the number of stereo pairs needed to cover all triangles in the target tetrahedron model, and also ensures that the target object is covered by at least one stereo pair:

$$\min \sum_{i=1}^{K_s} S_i , \tag{4.4}$$

where the S_i is a binary variable where a "1" indicates the chosen stereo pair. The function is subjected to the constraints of the surveillance model and the constraint that ensures that there is only one camera at each position. K_s is the total number of potential

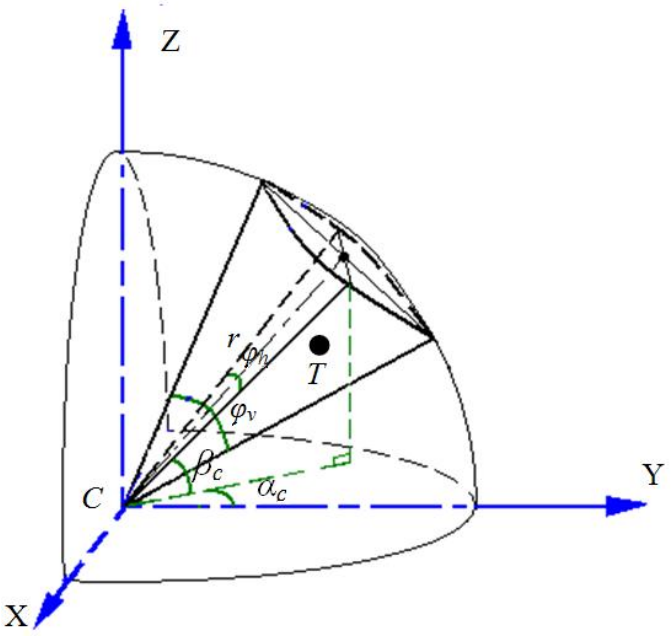


Figure 4.2. The spherical coordinate system and FoV of a camera where C is the camera position and the example target point is located at the point T.

stereo cameras.

Planning the placement of the stereo pair placement with the help of the integer linear programme consists of three stages:

- First, with the aid of the greedy algorithm, we find the potential stereo pairs that satisfy the stereo constraints from all potential cameras' positions and orientations, as presented in Paper V.
- Second, integer linear programming is applied to minimise the total number of stereo pairs subjected to the visibility and baseline length constraints, the depth accuracy constraints, and the camera movement distance constraints. The objective function minimises the number of stereo pairs needed to cover all triangles in the target object model and also ensures that the target object is covered by at least one stereo pair.
- Finally, the 3D reconstruction's accuracy can be verified by a cubic reconstruction.

4.2 Tracking algorithms and their implementations

For surveillance, the position of the target has to be predicted by a tracking algorithm that allows the system to keep the human in the camera's FoV. Thus, the surveillance algorithm can arrange the vision sensors' position and determine their orientation, movement speed, focal length and baseline length, [54], [55].

Rincón proposes that human tracking can be divided into three levels: the *punctual, region and human pose levels* based on the model of the target, [56]. The first level considers the human as a compact and small target, the second level extends the human as a region which can be defined by a human body and the third level models the human as a non-rigid object. In our approach, a compact model is used and denoted as a point that also represents the human position and the centre of gravity of the tetrahedron model.

In single-target tracking, only one target may be detected by the sensor. The approach usually assumes that a linear and Gaussian dynamical measurement model is used. The posterior density function gives all observations and can be computed from the Bayes recursion. For example, single tracking can be conducted with the aid of a standard Kalman filter, based on the linear dynamical systems, [57]. The extended Kalman filter based on the non-linear model can also be used for single-target tracking, [58].

In multi-target tracking, the number of targets is unknown and can vary with time. When the number of targets changes due to targets appearing and disappearing, tracking becomes complicated since it becomes difficult to obtain the posterior density function. In this case, a Probability Hypothesis Density, PHD, gives a solution and recovers the posterior density function approximately, [59]. Most traditional multi-target tracking algorithms such as Multiple Hypothesis Tracking, MHT, [60] and the Joint Probabilistic Data Association Filter, JPDAF, [61] involve explicit association between measurements and targets. Recently, there has been increasing interest in using random set theory to solve multi-target tracking. The multiple-target recursive Bayes filter,

based on the random Finite Set Statistics, FISST, is theoretically the optimal approach to multiple-sensor multiple-target detection, tracking, and identification, [62].

The GM-PHD algorithm provides a closed form solution to the PHD filter suitable to track multi-target movement assuming that the system is a linear Gaussian system, [43]. The measurement of multiple targets in the tracking system can be modelled by Random Finite Sets, RFS. The appearance of new objects can be described as RFS of spontaneous births (i.e. independent of any existing target) or spawning from an existing target. The GM-PHD algorithm estimates the number of targets and their states at each point in time. The algorithm is a recursion consisting of two stages: prediction and update.

The prediction stage estimates and produces a hypothesis about the new number and state of targets at the current time based on previous stages and can be calculated using the PHD. The PHD prediction stage depends on the intensity of the birth random finite sets, the intensity due to existing targets and the intensity of the spawned target. The expected number of targets can be estimated from the integration of the predicted intensity over all surveillance regions. The corresponding target states can be found from the peaks of the prediction intensity.

The update stage introduces the measurement information at the current time step to refine the previous prediction. In this manner, the object detection has been finished, the set of measurements of the multiple targets' positions is available, and the state is updated during this step. The update consists of two terms: the intensities of the mis-detected targets and of the detected targets.

The implementations of the GM-PHD algorithm using RFID and vision sensors are presented in Paper IV and VII respectively. The human tracking paradigm in the RFID and vision sensor systems is shown in Figure 4.3. The tracking filter is based on prior knowledge and observation. The observation can come from two systems separately. In the RFID system, the target's 3D position is calculated by the Levenber-Marquardt, LM, algorithm by using the RSSs. In the vision system, the target's position is calculated by the 3D reconstruction algorithm. The vision tracking system gives a relatively accurate human position prediction, but it requires an initial human position and is also limited by the light condition, by obstacles and by a narrow camera FoV. Here, the coarsely estimated target position produced by RFID can be used by a vision sensor to identify the region of interest or the targets' initial position. In the case when a vision system is disabled or limited, the target can still be tracked by an RFID sensor system.

In RFID systems, the target tag RSS is extracted from the readers. This depends on the distance between the tag, the reader's antennas and the propagation environment. From the received tag's RSSs, the distances from a tag to the readers can be calculated. In practice, the calculated distances are contaminated by noise so the optimization algorithm needs to be applied. The steps of implementation can be described as:

- First, the state of the multiple targets' positions can be measured by using the LM algorithm from the RSS measured by the RFID readers.
- Secondly, the target position is predicted by the GM-PHD algorithm.

In the stereo vision system, the target position can be estimated using parallel stereo vision sensors and the methods proposed in Chapter 3, such as triangulation or the dithering algorithm. The implementation of the GM-PHD algorithm in the vision sensor system can be divided into the following steps:

- First we obtain the 3D reconstructed target position.
- Second, each target is assigned by a fixed label in order to maintain target tracking continuation in the GM-PHD algorithm.
- Finally, the target position is predicted by the GM-PHD algorithm and the target position is measured by the stereo vision sensors.

In the RFID system, the target is identified by means of the tag's ID. This also makes it possible to assign the target a label in the vision system. The combination of the human tracking algorithm using RFID technology and the vision sensor system makes the IMSS more stable and robust.

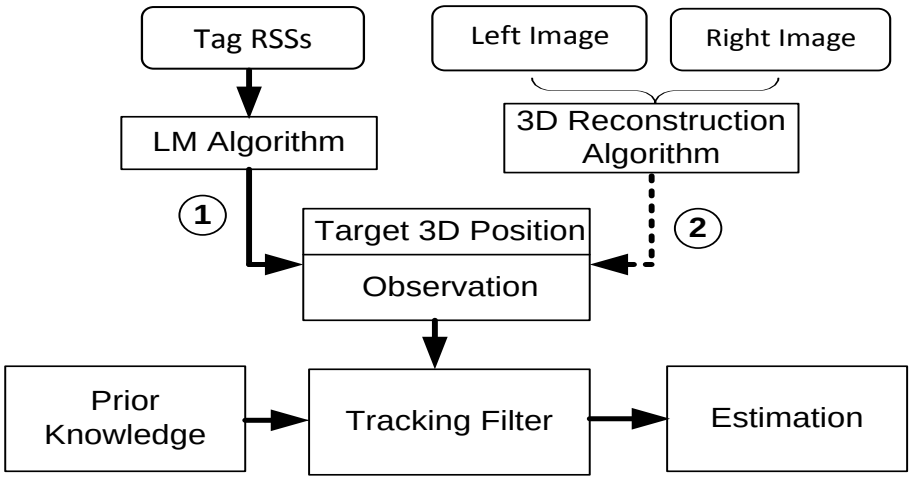


Figure 4.3. The human tracking paradigm in the RFID and vision system.

5 Validation and evaluation methods

Validation and evaluation are the processes used to confirm and/or assess a model or method. To reliably validate the measurement method in relation to both human perception and technological issues, it is important to include not only technical, but also human and environmental aspects of the model or process. In this chapter, we present two validation and evaluation methods for quantitative processes models, and one evaluation method for a hybrid process model that includes quantitative and qualitative factors. The validation method for 3D reconstruction presented in Paper II and III, and the evaluation method of the tracking performance described in Paper VII, are examples of quantitative methods. A hybrid evaluation method for image quality that combines human factors into the quality assessment is proposed in Paper VIII.

To validate a 3D reconstruction method that improves depth measurement accuracy is a complex process. The process can be simplified by applying a relative distance measurement instead of an absolute depth measurement. The relative distance, that is the information about the orders and/or ratios among objects' distances in relation to human vision depth perception, has been studied in the field of psychophysics. Ohtsuk et al. have reported that the integration of relative distance cues into the other depth perception cues, such as occlusion, convergence and aerial perspective, can improve absolute distance perception, [63]. The measurement of the differential depth by two un-calibrated cameras has been studied by Searle and Green, [64]. The idea in the verification part of Papers II and III is to validate the depth reconstruction method by measuring the differential depth.

Performance evaluation is an important issue in the design of a tracking algorithm. Juang and Burlina [65] propose a performance evaluation of the GM-PHD filter used in multi-target tracking by applying different levels of noise. The performance evaluation of multi-target tracking based on the Wasserstein distance has been studied by Clark et al., [43]. Furthermore, an evaluation of the multi-target tracking filter by applying a new Optimal Sub-Pattern Assignment, OSPA, metric in the context of multiple-target miss-distances has been introduced by Schuhmacher et al., [66]. In order to investigate how the characteristics of human motion, such as speed and angular velocity, affect the tracking filter performance, an evaluation of the GM-PHD filter tracking performance is introduced in Paper VII. The advantage of the circular motion model is that the target motion characteristics, in terms of speed and angular velocity, can be easily related to the radius and measurement sampling rate.

When evaluating image quality, several multidimensional aspects need to be considered. There are different image quality indices, depending on the application area. Wang et al propose a universal image quality index that is based on the difference

between a reference image and the measured image. It has been indicated that the index correlates with the human visual system and thus with human assessment, [20], [21].

5.1 Validation method of the depth reconstruction

The accuracy of the depth reconstruction depends on many factors, such as the stereo baseline length, the focal length and the camera sensor resolution. It is crucial to find a method that reduces the depth reconstruction uncertainty, e.g. the method presented in Paper II and III. To validate this novel 3D reconstruction method, we propose to measure the differential depth between two target points *A* and *B* instead of measuring the absolute distance.

The reconstructed quantised depth of a target point is the distance from the baseline to the target point computed from the discrete disparity found in a stereo image. The quantised differential depth, Z_{AB}, between two target points *A* and *B* from the stereo images can be measured from the difference between the quantised depths Z_A and Z_B of the target points *A* and *B* respectively. The Probability Distribution Function, PDF, of the differential depth quantisation uncertainty can be described as the convolution of the depth quantisation uncertainties ΔZ_A and ΔZ_B of the two target points *A* and *B* respectively:

$$p(\Delta Z_{AB}) = p(\Delta Z_A) \otimes p(\Delta Z_B), \tag{5.1}$$

where $\otimes$ denotes convolution.

The PDF of the depth quantised reconstruction uncertainty ΔZ, corresponding to an interval between two iso-disparity lines, is defined as [67]:

$$p(\Delta Z, n_t) = \begin{cases} \dfrac{n_t^2 \Delta D}{(bf)^2}\left(\Delta Z n_t^2 \Delta D + bf\right), & -\Delta Z_m \le \Delta Z < 0 \\ -\dfrac{n_t^2 \Delta D}{(bf)^2}\left(\Delta Z n_t^2 \Delta D - bf\right), & 0 \le \Delta Z \le \Delta Z_m \\ 0, & \text{elsewhere} \end{cases}, \tag{5.2}$$

where ΔD is the pixel size and n_t is an integer number representing the target disparity. ΔZ_m is the maximum depth reconstruction uncertainty corresponding to the interval between the iso-disparity surfaces. The focal lengths of the two cameras are assumed to be the same and are denoted as *f*. The stereo camera baseline is denoted as *b*.

Figure 5.1 illustrates plots from equations (5.1) and (5.2). Here, the baseline length *b* is 100 mm, the pixel size ΔD = 12.9 µm, and the focal length *f* = 25 mm. Figure 5.1(a) shows the normalized PDF of the depth quantisation uncertainty of the target point *A* located at a distance corresponding to the iso-disparity plane of number 100, and is taken as a reference point in the following three study cases. These three study cases consider three target points, *B*, B_1 and B_2, which correspond to the iso-disparity line numbers of 110, 150 and 400, respectively.

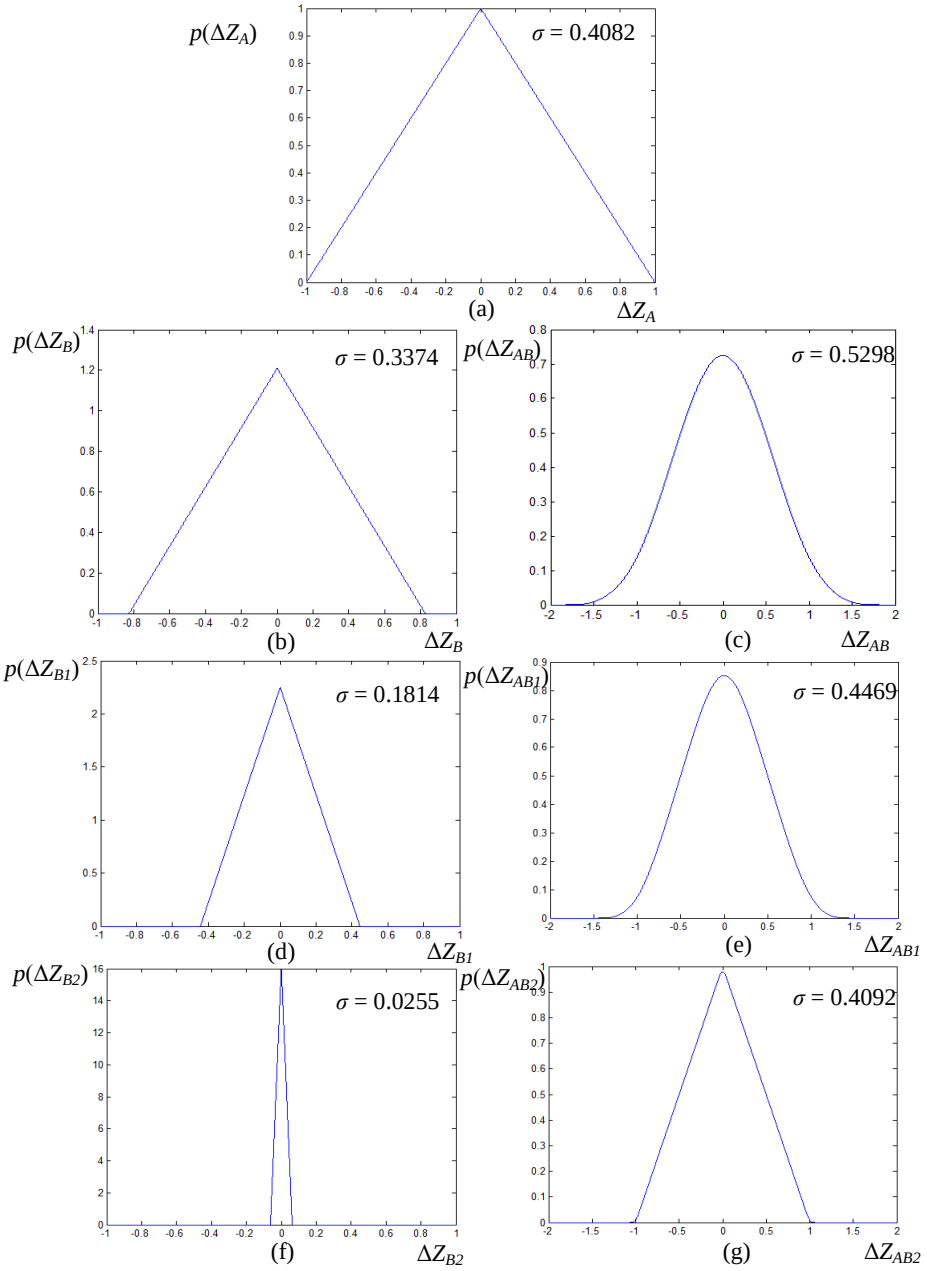
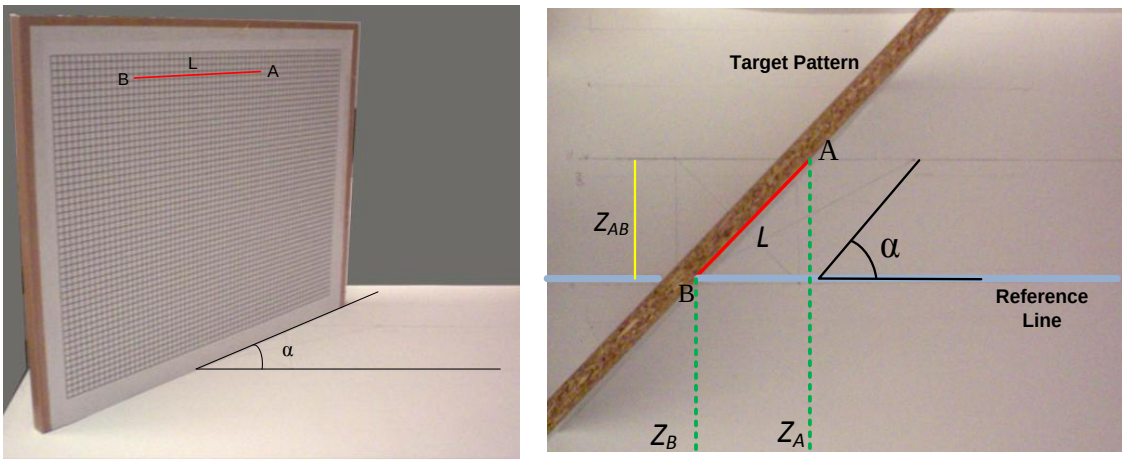


Figure 5.1. The PDF of the depth reconstruction uncertainty, (a) for point A located at the iso-disparity plane with the number 100 (the depth Z_A = 1.94 m), and (b), (d), (f) for the points B, B_1 and B_2 located at the iso-disparity plane with the numbers 110, 150 and 400, respectively (Z_B = 1.76 m, Z_{B1} = 1.29 m, and Z_{B2} = 0.48 m). The PDFs of the differential depth uncertainty between the point A and the points B, B_1 and B_2 are illustrated in (c), (e) and (g) respectively.

The differential depth quantisation uncertainties for the three study cases are presented in Figure 5.1 (c), (e) and (g), respectively. The uncertainty range is normalized by the depth quantisation uncertainty at the reference point *A* reconstructed with the greatest quantisation error. Figure 5.1 (b), (d) and (f) are plots of the normalized PDFs of the depth quantisation uncertainty for the three target points *B*, B_1 and B_2 respectively. The differential depth is the distance between two target points perpendicular to the baseline. The standard deviation σ of the uncertainty for each case



(a) (b)

Figure 5.2. The target grid position: (a) front view, (b) top view.

is indicated in the right corner of each figure. The range of the differential depth quantisation uncertainty is the sum of the depth quantisation uncertainty ranges of the two corresponding points.

The validation of the 3D reconstruction method accuracy is based on a comparison of the measured quantised differential depth with a reference differential depth taken from a grid pattern, e.g. as shown in Figure 5.2(a). If the target grid pattern is placed parallel to the baseline, points *A* and *B* will have a zero differential depth. However, by tilting the pattern with the angle α, a different reference differential depth is obtained. The tilting angle, α, is adjusted with respect to a reference line that is parallel to the baseline of the stereo system. If the distance between the two points on the grid, *L*, and the angle α are known, the reference differential depth Z_{ABr} can be directly obtained by: $Z_{ABr} = L \times \sin\alpha$.

The measurement of the differential depth is relatively simple and accurate, and can be used to validate the depth reconstruction methods. This validation method allows us to get rid of the calibration problem and avoids the problem of measuring the distance between the target and the centre of the lens.

5.2 A 3D spatial motion model for the tracking performance evaluation

The 3D circular motion characteristics of the target in terms of speed and angular velocity can be used for the evaluation of the tracking filter performance. The performance is measured in terms of a mean absolute error defined as the difference between the prediction and the ground truth by means of the Wasserstein distance, [43].

Figure 5.3 illustrates the 3D circular motion test signal, where the target point $\zeta_{k\text{-}1,\,j}$ at the time *k*-1 represents the target *j* position in the 3D space. This target is assumed to have moved to a new position $\zeta_{k,\,j}$ at the time *k*. The target test motion trajectory can be described as a circle with the centre *O* and the radius *r*. It is characterised by the direction vector $\alpha_{k\text{-}1,j}$ of the velocity of the target *j* in the 3-D space. The angle $\theta_{k\text{-}1,j}$ is the angle between the motion direction and the *YZ* plane, and the angle $\varphi_{k\text{-}1,j}$ is the angle between the motion direction and the *XY* plane. Vector $V_{k\text{-}1,j}\alpha_{k\text{-}1,j}$ denotes the velocity

during the time *k*-1, where $V_{k\text{-}1,j}$ is the speed (the absolute amplitude of velocity). The target position $\zeta_{k,j}$ during the circular motion can be approximated by:

$$\zeta_{k,j} = \zeta_{k-1,j} + r\omega\Delta k\left(\cos\theta_{k-1,j}, \sin\theta_{k-1,j}, \sin\varphi_{k-1,j}\right) \quad , \tag{5.3}$$

where ω is the angular velocity, r is the circular radius and Δk is the sample interval.

The target moves along the 3D spatial motion circle. The target motion speed is constant at any point during the circular motion for the same radius and angular velocity. The test motion speed can be described as $r\omega$. The angular shift per sample interval is $\Delta\theta = 360\,^{\circ}/K$, where K is a sampling rate. The ability of the filter to track targets with increasing motion speed or angular velocity can be evaluated by changing the radius r or the sampling rate K.

In order to evaluate the multi-tracking continuity, two targets whose trajectories follow the 3D circular motion are proposed. The initial position $\zeta_{0,j}$ of target j in the 3D space is shown in Figure 5.3 and described by:

$$\zeta_{0,j} = O + r_j\left(\cos\alpha_{0,j}, \sin\alpha_{0,j}, \sin\beta_{0,j}\right) \quad , \tag{5.4}$$

where $O=(O_x, O_y, O_z)$ is the circle centre, $\alpha_{0,j}$ is an angle between the radius vector and the YZ plane and $\beta_{0,j}$ is the angle between the radius vector and the XY plane. By choosing the different initial angle to $\alpha_{0,j}$ and $\beta_{0,j}$, the two targets may approach each other in the X, Y and Z dimensions, respectively. This causes occlusions in each dimension, respectively. The two targets' motion trajectories are shown in Figure 5.4. Each target need to be assigned by a label. The two targets' labels should be captured throughout the whole tracking process. The label continuity maintained by the filter can be verified when the occlusion happens.

5.3 The image quality evaluation method

The image quality evaluation method is an adaptive method consisting of both quantitative and qualitative factors. The neural network is a suitable tool to integrate these two factors. Image quality is characterized by quantitative features such as basic properties (resolution, luminance, contrast and size), naturalness (sharpness, structure and noisiness), and colourfulness (colour temperature, chroma and saturation). However, the human assessment of the image quality depends also on many qualitative factors such as personal background, physical environment, usefulness, tools, and pattern representation, all related both to the target and to the human being.

Figure 5.5 illustrates the evaluation method of a quality index for a general product, service or condition. The initial quality index is established by experts in the field. The quantitative features' initial weights, [α], used in the measurement of the quality index, are based on the *measurement uncertainty* and *relevance* of each *feature*. Then, the *adaptive measurement method* applies a training process of integrating the relationship between the value of the *quantitative features* and the *subjective human assessments* regarding quality.

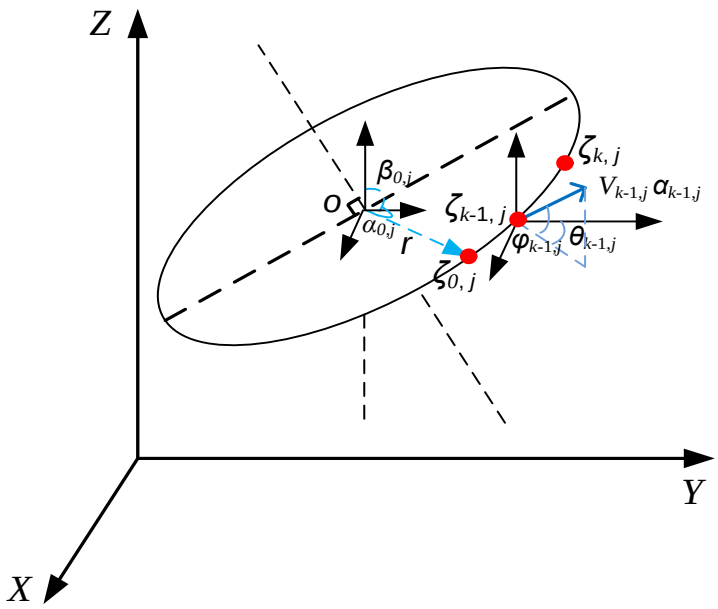


Figure 5.3. The illustration of the circular motion of the target

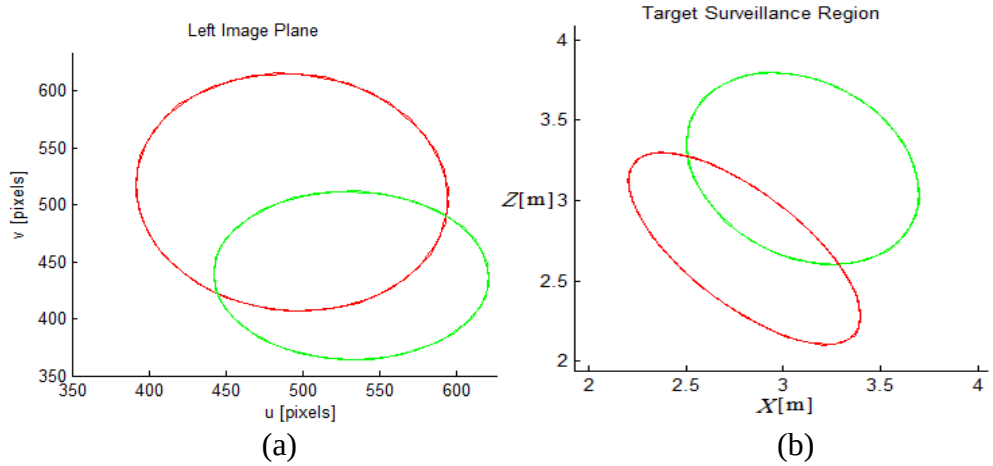


Figure 5.4. The motion trajectories of two targets represented by red and green curves respectively. (a) Left image plane view and (b) top view of the scene.

The adaptive image quality evaluation method can be divided into three steps:

- First, we chose to classify the images according to different groups of people. The groups' assessments can be biased due to gender and/or because they may have had previous experience with image processing. The reconstructed quantified assessments are computed from the first three principle components resulting from the application of Principle Components Analysis, PCA. Next, the mean values of the reconstructed grades were taken for each image within each group.
- Second, the adaptive quality model based on the neural network is applied. Three types of quantitative features are used during the training stage: *structure distortion ratio* along with the two basic properties *luminance*

distortion ratio and *contrast distortion ratio*. The model is trained on the same image but with different disturbances.

- Finally, after training the evaluation stage occurs. In our study, the quality index evaluated by the neural network and the ranking produced by the human judgments matched each other very well. Thus, based on the result from Paper VIII, one can conclude that the model recognises different kinds of disturbance.

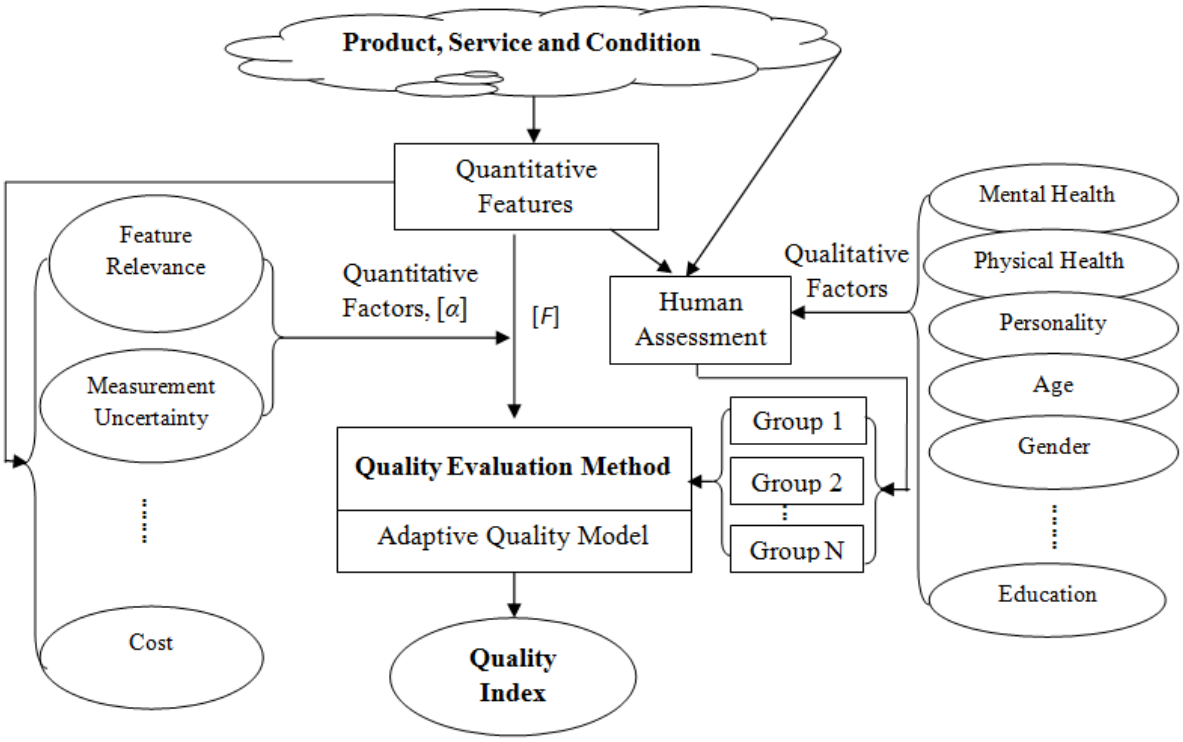


Figure 5.5. A block diagram illustrating the quality evaluation method. Ellipses denote representations of information, and rectangles denote process transformations from one representation into another.

6 Summary

6.1 Overview of the papers

This chapter gives a brief summary of the eight papers included in the thesis. The papers are related to three aspects: depth measurement, tracking and surveillance, and evaluation and validation as described in Figure 1.4. The summary describes how the studies were performed, the results obtained, and the conclusions drawn.

6.1.1 Paper I - Planning of a multiple sensor system for a human activities space – aspects of iso-disparity surface

For a stereo pair of sensors, the 3D reconstructed space is quantised by the iso-disparity surfaces and the depth reconstruction accuracy is defined by the intervals between the iso-disparity surfaces. A mathematical geometry model is used to analyse the iso-disparity surfaces. This model is the fundamental theory in depth reconstruction uncertainty analysis and used in camera arrangements for tracking and surveillance. The model can be used to dynamically adjust the positions, orientations and baseline lengths of multiple stereo pairs of cameras in a 3D space in order to get sufficient visibility and accuracy for surveillance, tracking, and 3D reconstruction. The iso-disparity surface is a function of the baseline length, the focal lengths and the sensor pixel size for a general parallel stereo pair with zooming. For a general convergent stereo pair, the iso-disparity surface is also the function of the convergence angle. The depth reconstruction accuracy is quantitatively analysed by the proposed model.

The presented analysis shows that the depth reconstruction accuracy varies more significantly with respect to the target distance to the baseline, the baseline length, and the focal length than in relation to the convergence angle. Small changes in the stereo convergence angle do not affect the depth accuracy to any great degree, especially not when the target is placed centrally. On the other hand, the convergence angle can have a great impact on the shape of the iso-disparity curves and the size of the common FoV.

The proposed mathematical iso-disparity model makes it possible to perform a reliable control of the iso-disparity curves' shapes and intervals by applying the system's configuration and target properties.

6.1.2 Paper II - Depth reconstruction uncertainty analysis and improvement – the dithering approach

The depth reconstruction uncertainty is the main topic of this paper and can be analysed with the help of the iso-disparity surfaces. The paper describes the image quantisation uncertainty model and gives out the distribution function of the disparity quantisation uncertainty. The probability density function of the disparity quantisation uncertainty is

a triangular distribution. This is a result of the convolution of two rectangular distributions of the probability density functions of the left and right images' horizontal quantisation uncertainty. Furthermore, a mathematical model of the depth reconstruction uncertainty used for analysis of the reconstruction uncertainty is presented in the paper. A dithering algorithm is implemented to reduce the depth reconstruction uncertainty and the algorithm's theoretical background gives further guidance for the generation of a dither signal. The discrete dither signals are estimated by analysing the iso-disparity planes and then generated by controlling the stereo pair baseline length and moving the new iso-disparity plane into the exact middle of the previous iso-disparity planes. This gives optimal control of the camera movement in respect to the quantisation uncertainty improvement. By applying a two-stage discrete dithering signal and combining four images into four pairs of stereo images, the depth of the target point can be estimated without bias.

In the paper, the presented model is also applied to the identification of a required 3D reconstruction space with defined accuracy. This application extends a target point into a more realistic space.

The simulated statistical analysis of the depth reconstruction uncertainty reveals an improvement of the depth reconstruction accuracy by 49.7%. The physical experiment shows an improvement of the depth reconstruction accuracy by 36.2%. The differences in results between the simulation and the physical experiment can be attributed to other factors that the influence of the measurement. The results furthermore reveal that the target-depth reconstruction uncertainty is reduced by half by the proposed algorithm.

6.1.3 Paper III - Applying dithering to improve depth measurement using a sensor-shifted stereo camera

This paper introduces the use of sensor-shifted cameras in the stereo system and applies the dithering algorithm to improve depth reconstruction. The sensor-shifted stereo cameras give a large stereo common FoV for depth measurement, and do not require stereo rectification during matching process which is normally required for convergent stereo cameras. One more advantage is that this type of stereo cameras can reduce the motion blur when using the dithering algorithm.

A mathematical model of the depth measurement using the sensor-shifted parallel stereo system was developed. The model gives a function of the depth measurement with respect to the sensor shift distance, the baseline length, and the focal length and disparity. Furthermore, a dithering algorithm based on the depth measurement model is presented in the paper. The dithering signal is generated by the sensor chip shift. Both the synthetic and the physical experiments verify that applying dithering reduces the depth reconstruction uncertainty by half compared to the direct method.

It is verified by simulation that the dither signal causes the iso-disparity surfaces to shift in respect to their initial positions. The new position is in the middle of the iso-disparity intervals before applying the dither signal. This can be interpreted to mean that dithering can be applied to reduce the depth reconstruction uncertainty by half through a multiple-stage measurement.

In the synthetic experiment, the enhancement was verified using computer simulation. The depth reconstruction of the test target points using the dithering algorithm shows improvement of the uncertainty since the reconstruction of the points

forms iso-disparity lines with interval widths reduced by half for the dithering algorithm when compared to the direct method. This improvement is confirmed by comparing the histograms of the reconstruction error produced by both methods. This comparison shows that the dithering algorithm reduces the span of the error distribution by half when compared to the range obtained with the direct method. The depth reconstruction improvement can also be observed in the 48.6% reduction in the standard deviation of the reconstruction error for the simulation targets by the dithering algorithm.

The results of the physical experiment also show reconstruction improvement for all test targets with an average of 49%, which is close to the theoretical value. For higher experimental accuracy, the differential depth of the target pairs was used in the validation experiment instead of the absolute depth of a single target.

6.1.4 Paper IV - RFID multi-target tracking using the probability hypothesis density algorithm for a health care application

The investigation of multi-target tracking by an RFID system is proposed in this paper. The system can still locate the multiple targets' positions when the vision system is on standby or the targets are obscured by obstacles. The tracking performance improvement is accomplished by the application of the LM algorithm with the GM-PHD filter.

An optimal algorithm to estimate the propagation factor in the radio propagation channel model is first proposed in the paper. This provides the flexibility and robustness needed to adjust the propagation factor to different environments. Secondly, the tagged target position is estimated by the LM algorithm based on geometrical triangulation calculation using the signal strength from readers. Finally, the GM-PHD filter is implemented to track the target position which is estimated by the LM algorithm.

The ability of the GM-PHD algorithm for multi-target tracking is validated by simulation and by real environment experiments. It is shown that the algorithm successfully tracks multiple targets while using the RSSs of each tag. By using the LM algorithm with the GM-PHD filter, the simulation and physical experiments show an improvement of the mean value by 33% and 29%, respectively, and an improvement of the standard deviation of 37% and 33%, respectively, when compared to using the LM algorithm without the GM-PHD filter.

6.1.5 Paper V - Planning of a multi stereo visual sensor system for a human activities space

In this paper, in order to get efficient visibility for surveillance, tracking, and 3D reconstruction, a new approach to optimise the multiple stereo visual sensor configurations is discussed. The optimisation is implemented by applying the camera, the object, and the stereo pair constraints into the integer linear programming.

The camera's 3D field of view is modelled by spherical coordinates, which speeds up computation. The human target space is modelled as a tetrahedron. This model allows for convenient extraction of the orientation of each surface, which in turn guarantees good observability. The stereo pairs can be formulated by making use of the greedy algorithm using stereo constraints to acquire all possible stereo pairs. By analysis of the constraints, the minimum number of stereo pairs necessary to cover the

entire target space and the camera pairs' orientations are optimised by integer linear programming.

The presented simulations were performed in order to obtain the optimal number of stereo pairs along with the corresponding camera positions and orientations according to the target location and the required constraints. The simulations proved that a set of two pairs is sufficient to observe the human modelled as a tetrahedron, on the condition that in each position all upward triangle surfaces are visible to at least one stereo pair.

6.1.6 Paper VI - Planning of a multi stereo visual sensor system - depth accuracy and variable baseline approach

In this paper, the key factors that affect the accuracy of 3D reconstruction are analysed. The paper argues that the convergence angle and the target distance significantly influence the depth reconstruction accuracy. The depth accuracy constraint is implemented in the model to control the stereo pair's baseline length, position, and orientation. The depth accuracy constraint guarantees a certain accuracy in the 3D reconstruction. The reconstruction accuracy is verified by a cubic reconstruction method.

The simulation results show that the cubic reconstruction method is useful when verifying the reconstruction accuracy and essentially proves that the proposed method of controlling the baseline length is functional. In order to follow the movement of the target object, the camera's movement distance constraint is applied in the optimisation programme, and two-stage camera sampling is implemented. This two-stage camera position sampling allows for flexible adjustment of the position ranges and intervals and thus speeds up computation.

6.1.7 Paper VII - Evaluation of the GM-PHD filter for multi-target tracking with a stereo vision system

The human movement property is one of the most influential factors in the filter-tracking performance evaluation. The circular 3D spatial motion is modelled in a 3D space by two key parameters, speed and angular velocity, which play key roles in human movement and are useful for the evaluation of the filter tracking performance. The advantage of the circular motion model is that the parameters of the target motion and the measurement can be easily related to the radius and the measurement sampling rate. This model is introduced as a test signal for the evaluation of the accuracy and label continuity of the filter.

The GM-PHD filter for multi-target tracking is modelled and implemented in MATLAB and validated by applying two random motion trajectories. The robustness of the filter performance when occlusions occur during multi-target tracking is also investigated in a simulated environment by using the human circular motion model. This makes it possible to handle occlusion by gathering 3-D information from the activities space. It is shown that the filter successfully tracks targets while maintaining each target's label and performs well even when one target is occluded by another one, or when they cross each other.

The mean absolute error between the prediction and the ground truth in the tracking shows that the GM-PHD filter performance depends on target motion speed and angular

velocity, and that the tracking error is thus proportional to the motion speed and angular velocity.

6.1.8 Paper VIII - An adaptive quality assessment system – aspect of human factor and measurement uncertainty

This paper proposes a model of an adaptive system for image quality measurement. The system can handle both qualitative and quantitative factors that are a part of the image quality index. Furthermore, the proposed objective group classification method is useful in cases when the quality assessment of different customers/user groups differs significantly. As a modelling tool, the neural network is used. With the help of the neural network, the system integrates the human qualitative judgement with quantitative measurements to create a quantitative index.

The experiment results presented in the paper show that the quality index, QI, estimated by the adaptive system and the human quantitative assessments matched each other very well.

6.2 Conclusions

To assure that the proposed Intelligent Multi-Sensor System, IMSS, which is a high-performance autonomous distributed vision and information processing system, is suitable for surveillance of a human activities space, many research problems, several of which have been addressed in this thesis, must be resolved.

The depth reconstruction uncertainty problem has been approached from the point of view of modelling, improvement methods, implementation and validation by simulations and by real experiments. The proposed iso-disparity surfaces model simplifies an evaluation of the measurement methods based on depth reconstruction. This model facilitates a quantitative analysis of the depth reconstruction uncertainty and can also be useful for improving the reconstruction uncertainty by controlling the multiple stereo pairs' positions, orientations, and baselines. Due to its robustness, the iso-disparity surfaces model can be applied when surveying a moving target.

Signal processing methods can also help to reduce the depth measurement uncertainty. For instance, the dithering algorithm is a suitable method to improve the depth reconstruction accuracy. Fusing a few images taken by low-resolution sensors leads to high resolution in the depth reconstruction. It has been proved that the depth reconstruction uncertainty can be reduced by half by the dithering approach when compared to the direct triangulation method. The dither signal is controlled by means of the iso-disparity planes based on the stereo pair's baseline length or sensor chip shift.

The proposed dither signal generated by the stereo setup with a stationary lens and a micro-shifted sensor improves the accuracy of the 3D reconstruction and also reduces the effect of motion blur when using the sensor-shifted stereo camera.

The GM-PHD algorithm implemented in the vision-sensor and in the RFID sensor systems gives reliable and continual tracking of multiple targets. In the vision-sensor system, the GM-PHD algorithm uses stereo images to track the target's 3D position. In the RFID system, the GM-PHD algorithm successfully tracks multiple targets while using the RSSs. The GM-PHD algorithm estimates the target's position more accurately than the LM algorithm.

The proposed surveillance algorithm implemented in a controllable multiple stereo vision system controlled by the tracking signal guarantees that the target is observed under efficient visibility and with the required depth reconstruction accuracy. Also, application of the target space constraints and the stereo pairs' properties improve the surveillance accuracy. Here, integer linear programming is useful to minimise the number of stereo pairs necessary to cover the entire target space under efficient visibility and the required depth reconstruction accuracy. Furthermore, modelling the human being as a tetrahedron is a convenient way to extract the orientation of each surface and guarantees good observability.

The evaluation and validation methods applied to depth measurement, image quality and tracking performance facilitate the comparative analysis of the proposed methods and their improvements. The differential depth measurement between two target points is recommended to verify the accurate improvement of this novel depth reconstruction method. This validation method overcomes the difficulty of making an accurate measurement of the absolute distance from a camera to the target. The measurement of the differential depth is relatively simple to perform and gets rid of the calibration problem.

The evaluation of the image quality by an adaptive method is introduced. It makes use of the FDV which integrates the human factor into the evaluation of the image quality. The model focuses on the human assessment of image quality. This is useful when there is a lack of a clearly defined quantity or when there is no physical standard to be used in calibration. The weights of each quantitative feature from the image, such as the basic properties, the naturalness and the colourfulness, should be based on the measurement uncertainty and on its relevance for quality assessment. The proposed image quality assessment using an adaptive model integrates both quantitative and qualitative factors. As an implementation tool, the neural network has been used.

The GM-PHD filter performance for multi-target tracking can be evaluated by measuring a mean absolute error in the Wasserstein distance. The target movement features such as speed and angular velocity play a key role in the evaluation of the tracking algorithm performance. The evaluation can be accomplished by using a 3D spatial circular movement as the test signal. For the circular motion model, the target motion and its measure can be easily defined by the test signal parameters such as radius and sampling rate. The quantitative analysis of the performance of the tracking algorithm implemented in the stereo vision system is done by means of the absolute error versus the target motion speed and the angular velocity. This proves that the proposed test signal is suitable for performance evaluation.

6.3 Future research

The work presented in the thesis constitutes a first step towards advanced IMSS design. There are many scientific problems and questions which have to be explored and further developed to realise the advanced IMSS. Future work in this area can be categorized into two branches: extension of IMSS applications and functions, and development of new techniques and algorithms that improve IMSS performance.

It would be interesting and important to investigate the implementation possibility of the reconstruction method for specific health care applications and motion rehabilitation. Here, it may be possible to construct a 3D vision system able to support

monitoring and subsequent correction of human motions when patients perform their rehabilitative training. The analysis results can be used by physical therapists for further diagnosis and rehabilitation planning. Another approach which can attract particular attention is an ambulatory system used by elderly individuals in their own homes and designed to detect abnormalities in their motions and to report these abnormalities to a remote monitoring centre for further action. Such a system could even be implemented for monitoring patients in health care institutes.

The idea of an adaptive quality model presented in Paper VIII is a promising starting point for the introduction of the quantitative human factor into future vision systems. Such a system by be able to learn human habits or recognize human health/mood statuses and then apply this learning to the system's decision stage which can then be adapted to the personal character of the supported individual. As an example of an application of the adaptive measurement model in quality assessment, the possibility of an intelligent television can be explored. The IMSS can recognise, identify, and learn the human preferences regarding the television screen's colour, contrast, saturation, etc. According to this person's individual preferences, the system can then make decisions regarding peripherals, automatically controlling the television's settings and the surrounding environment including lights, curtains, etc. Another application example is how the adaptive system may find the proper stereo pair positions, orientations, and image processing algorithms to improve the image quality for health/mood monitoring and diagnostics applications. The adaptive measurement method related to quality can be applied to the 3D reconstruction accuracy strategy, which can then be adjusted according to both human and environmental factors.

The modelling of a human being for tracking purposes gives a large number of possible extensions of the research work. It is, for example, possible to use the human geometry model proposed in this thesis to distinguish the front or facial side of the human being from the back side, something that generally requires higher resolution in a 3D reconstruction.

The fusion of information from multi-sensor and multi-source systems is an interesting topic to study. The required information needs to be systematically and efficiently interpreted from a large volume of information which is acquired from different sensors and with different degrees of uncertainty. The fusion system can dynamically extract the relevant sensory data for a given purpose and within a specified time period.

The multi-target tracking algorithm used when the target is occluded for a long period of time needs further development. The usage of multiple cameras to solve occlusion problems in motion tracking that employs the GM-PHD filter seems to be a useful way to address this problem.

Future research could also focus on the integration of the vision and RFID systems. In particular, studies may focus on the way that the tag position measured by the vision system may be used as a reference to calibrate the RFID system to adjust the radio propagation factor according to changes in the environment.

During our research, it became apparent that the sensor-shifted micro movement caused by the dither signal affected not only the projection position of the target point but also changed the image pixel intensity of this point. This suggests that the projection

image pixel intensity change can be combined into the depth reconstruction using the dithering algorithm, thus achieving a more accurate depth reconstruction.

References

[1] MIT Project Oxygen, http://oxygen.csail.mit.edu, [Accessed 13 Jan., 2011].

[2] H. Hashimoto, "Intelligent space: interaction and intelligence", *Artificial Life and Robotics*, vol. 7, no. 3, pp. 79-85, 2003.

[3] Y. J. Kil, B. Mederos, and N. Amenta, "Laser scanner super-resolution", in: *Eurographics Symposium on Point-Based Graphics*, Massachusetts, USA, 2006.

[4] R. A. Wannamaker, *The Theory of Dithered Quantization*, PhD thesis, the University of Waterloo, Waterloo, Canada, 2003.

[5] R. A. Wannamaker, S. P. Lipshitz, J. Vanderkooy, and J. N. Wright, "A theory of nonsubtractive dither", *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 499-516, 2000.

[6] M. F. Wagdy, "Effect of various dither forms on quantization errors of ideal A/D converters", *IEEE Transactions on Instrumentation and Measurement*, vol. 38, no. 4, pp. 850-855, 1989.

[7] A. Mittal, "Generalized multi-sensor planning", in: *Proc. of 9th European Conference on Computer Vision*, Graz, Austria, 2006.

[8] X. Chen, *Design of Many-Camera Tracking Systems for Scalability and Efficient Resource Allocation*, PhD thesis, Stanford University, Palo Alto, USA, 2002.

[9] K. Chakrabarty, S. Iyengar, H. Qi, and E. Cho, "Grid coverage for surveillance and target location in distributed sensor networks", *IEEE Transaction on Computers*, vol. 51, no. 12, pp. 1448-1453, 2002.

[10] E. Hörster and R. Lienhart, "On the optimal placement of multiple visual sensors", in: *Proc. of ACM International Workshop on Video Surveillance & Sensor Networks*, California, USA, 2006.

[11] U. Erdem and S. Sclaroff, "Automated camera layout to satisfy task-specific and floor plan-specific coverage requirements", *Computer Vision and Image Understanding*, vol. 103, pp. 156-169, 2006.

[12] W., Yao, C. H. Chu, and Z., Li, "The use of RFID in healthcare: benefits and barriers", *IEEE International Conference on RFID-Technology and Applications*, Guangzhou, China, 2010

[13] C. J. Lin, T. L. Lee, S. L. Syu, and B. W. Chen, "Applications of intelligent agent and RFID technology for indoor position: safety of kindergarten as example", in: *Proceedings of the 9th International Conference on Machine Learning and Cybernetics*, Qingdao, China, 2010.

[14] J. M., Jorge, "The Levenberg-Marquardt algorithm: implementation and theory", *Numerical Analysis*, vol. 630, pp. 105-116, 1978.

[15] B. Vo and W. K. Ma, "The Gaussian mixture probability hypothesis density filter", *IEEE Transactions Signal Processing*, vol. 54, no. 11, pp. 4091-4104, 2006.

[16] J. Wirandi, *Modelling and Validation of Industrial Measurement Systems- Aspects of Quality and Human Factors*, PhD thesis, Lund University, Lund, Sweden, 2007.

[17] A. Bremner, "A convenient easy to use system for estimating the quality of chilled seafood", *Fish Processing Bulletin*, no. 7, pp. 59-73, 1985.

[18] N. Stambuck-Giljanovic, "Water quality evaluation by index in dalmatia", *Water Research*, vol. 33, no.16, pp. 3423-3440, 1999.

[19] P. G. Engeldrum, "Image quality modeling: where are we?", in: *Proc. IS & T's PICS Conference*, pp. 251-255, Savannah, Georgia, 1999.

[20] Z. Wang and A. C. Bovik, "A universal image quality index", *IEEE Trans. Signal Processing Letters*, vol. 9, no. 3, pp. 81-84, 2002.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.

[22] J. Wirandi, J. Chen, and W. Kulesza, "An adaptive model of the fuzzy variable – quality index", *AMUEM 2007 – International IEEE Workshop on Advanced Methods for Uncertainty Estimation in Measurement*, Trento, Italy, 2007.

[23] M. B. Miles, M, A. Huberman, *Qualitative Data Analysis*, 2nd ed., Sage Publication, Thousand Oaks, USA, 1994.

[24] M. Balnaves and P. Caputi, *Introduction to Quantitative Research Methods:An Investigative Approach*, SAGE Publications, London, UK, 2001.

[25] L. R. Gay, *Educational Research: Competencies for Analysis and Application*, 5th ed., Prentice-Hall Inc., Columbus, UK, 1996.

[26] F. C. Goldstein and H. S., Levin, "Disorders of reasoning and problem-solving abilities", in: A. L. Benton, M. R. Meier, & L. Diller (Eds.), *Neuropsychological Rehabilitation*, Guilford Press, New York, USA, 1987.

[27] E., Amsel, R., Langer, and L., Loutzenhiser, "Do lawyers reason differently from psychologists? A comparative design for studying expertise", in: R. J. Sternberg & P. A. Frensch (Eds.), *Complex Problem Solving: Principles and Mechanisms*, pp. 223-250, Lawrence Erlbaum Associates, Hillsdale, USA, 1991.

[28] R. E. Mayer, *Thinking, problem solving, cognition*, 2nd ed., W. H. Freeman and Company, New York, USA, 1992.

[29] I. L. S.; Sheng and T. Kok-Soo, "Eco-efficient product design using theory of inventive problem solving (TRIZ) principles", *American Journal of Applied Sciences*, vol. 7, no. 6, pp. 852-858, 2010.

[30] T. S. Kuhn, "The function of measurement in modern physical science", *Isis*, no. 52, pp. 161-193, 1961.

[31] J. McGhee, V. Henderson, M. J. Korczynski, and W. Kulesza, *Scientific Metrology*, 1st ed, Lodart S.A., Lodz, Poland, Reprint with corrections, 1996.

[32] M. Pollefeys, *Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences*, PhD thesis, Katholieke University, Leuven, Belgium, 1999.

[33] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, Cambridge, UK, 2004.

[34] B. Völpel and W. M. Theimer, "Localization uncertainty in area-based stereo algorithms", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, no. 12, pp. 1628-1634, 1995.

[35] M. Pollefeys and S. Sinha, "Iso-disparity surfaces for general stereo configurations", in: *Proc. of the 6th European Conf. on Computer Vision*, Dublin, Ireland, 2004.

[36] K. Ogle, *Researches in Binocular Vision*, W.B. Saunders Company, Philadelphia & London, UK, 1950.

[37] S. C. Park, M. K. Park, and M.G. Kang, "Super-resolution image reconstruction: a technical overview", *IEEE Signal Processing Magizine*, vol. 20, no. 3, pp. 21-36, 2003.

[38] P. Vandewalle, S. Süsstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution", *EURASIP Journal on Applied Signal Processing*, pp. 1-14, 2006.

[39] T. Chen, P. Catrysse, A. Gamal, and B. Wandell, "How small should pixel size be?", in: *Proc. of SPIE on Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, vol. 3965, 2000.

[40] A. Francisco and F. Bergholm, "On the importance of being asymmetric in stereopsis - or why we should use skewed parallel cameras", *International Journal of Computer Vision*, vol. 29, no. 3, pp. 181-202, 1998.

[41] J. Wirandi and A. Lauber, "Uncertainty and traceable calibration – how modern measurement concepts improve product quality in process industry", *Elsevier Measurement*, vol. 39, no. 7, pp. 612-620, 2006.

[42] J. Wirandi, W. Kulesza, and A. Lauber, "Human factor validation in an industrial measurement system", *Elsevier Measurement*, vol. 41, no. 7, pp. 705-718, 2008.

[43] D. Clark, K. Panta, and B.-N. Vo, "The GM-PHD filter multiple target tracker", *9th International Conference on Information Fusion*, Florence, Italy, 2006.

[44] D. C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley and Sons, New York, USA, 1996.

[45] J. O'Rourke, *Art Gallery Theorems and Algorithms*, Oxford University Press, New York, USA, 1987.

[46] W. Kulesza, J. Chen, and S. Khatibi, "Arrangement of a multi stereo visual sensor system for a human activities space", in: A. Bhatti (Ed.), *Stereo Vision*, pp. 153-172, InTech Education and Publishing, Vienna, Austria, 2008.

[47] E. Samson, D. Laurendeau, M. Parizeau, S. Comtois, J. Allan, and C. Gosselin, "The agile stereo pair for active vision", *Machine Vision and Application*, vol. 17, pp. 32-50, 2006.

[48] Y. Nakabo, T. Mukai, Y. Hattori, Y. Takeuchi, and N. Ohnishi, "Variable baseline stereo tracking vision system using high-speed linear slider", *IEEE Int. Conf. on Robotics and Automation*, Barcelona, Spain, 2005.

[49] N. T. Pham, W. Huang, and S. H. Ong "Probability hypothesis density approach for multi-camera multi-object tracking", *Lecture Notes in Computer Science*, pp. 875-884, 2007.

[50] N. T. Pham, W. Huang, and S. H. Ong "Tracking multiple objects using probability hypothesis density filter and color measurements", *IEEE International Conference on Multimedia and Expo*, Beijing, China, 2007.

[51] B. Abdelmoula, S. Horacio, and M. Mitsuji, "RFID indoor positioning based on probabilistic RFID map and kalman filtering", in: *Proceedings of 3rd IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, New York, USA, 2007.

[52] R. Bajaj, S. L. Ranaweera, and D. P. Agrawal, "GPS: location tracking technology", *E-Journal Computer*, vol. 35, no. 4, pp. 92-94, 2002.

[53] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 37, no. 6, pp. 1067-1080, 2007.

[54] J. Chen, O. E. Adebomi, O. S. Olusayo, and W. Kulesza, "The evaluation of the Gaussian mixture probability hypothesis density approach for multi-target tracking", *IEEE International Conference on Imaging Systems and Techniques*, Thessaloniki, Greece, 2010.

[55] J. Chen, *A multi sensor system for a human activities space – aspects of planning and quality measurement*, Licentiate Dissertation, Blekinge Institute of Technology, Karlskrona, Sweden, 2008.

[56] J. M. Rincón, *Feature-Based Human Tracking: From Coarse to Fine*, PhD Thesis, University of Zaragoza, Zaragoza, Spain, 2008.

[57] B. Ristic, S. Arulampalam, N. Gordon, *Beyond the Kalman Filter: Particle Filter for Tracking Application*, Artech House, London, UK, 2004.

[58] M. Karan and D. W. McMichael, "A multisensor single target tracking simulator: MUST", *Symposium on Data Fusion*, Adelaide, Australia, 1996.

[59] R. Mahler, "Multitarget bayes filtering via first order multi-target moments", *IEEE Trans. Aerosp. Electron. Syst*., vol. 39, no. 4, pp. 1152–1178, 2003.

[60] S. Blackman, "Multiple hypothesis tracking for multiple target tracking", *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 5-18, 2004.

[61] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, Academic Press Professional, San Diego, USA, 1988.

[62] H. J. Zhang, Z. L. Jing, and S. Q. Hu, "Tracks extraction of the probability hypothesis density filter for survival targets", *IEEE Conference Proceedings of the 27th Chinese Control Conference*, pp. 343-347, Kunming, China, 2008.

[63] S. Ohtsuka, H. Ujike, and S. Saida, "Relative distance cues contribute to scaling depth from motion parallax", *Perception and Psychophysics*, vol. 64, no. 3, pp. 405-414, 2002.

[64] G. Searle and R. Green, "Relative disparity using motion in stereo vision", *Conference of Image and Vision Computing*, Dunedin, New Zealand, 2005.

[65] R. Juang and P. Burlina, "Comparative performance evaluation of GM-PHD filter in clutter", *the 12th International Conference on Information Fusion*, Washington USA, 2009.

[66] D. Schuhmacher, B. T. Vo, and B. N. Vo, "A consistent metric for performance evaluation of multi-object filters", *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3347-3357, 2008.

[67] J. Chen, "The depth reconstruction accuracy in a stereo vision system", *Metrologia: dzis i jutro*, Gdansk University of Technology, Gdansk, Poland, 2009.

Part II

Paper I

Planning of a Multiple Sensor System for a Human Activities Space – Aspects of Iso-disparity Surface

Authors:

Jiandan Chen, Siamak Khatibi, Jenny Wirandi and Wlodek Kulesza

Reformatted version of paper originally published in:

Proceedings of SPIE on Optics and Photonics in Security and Defence, vol. 6739, Florence, Italy, September, 2007

Planning of a Multiple Sensor System for a Human Activities Space – Aspects of Iso-disparity Surface

Jiandan Chen, Siamak Khatibi, Jenny Wirandi and Wlodek Kulesza

Abstract

The Intelligent Vision Agent System, IVAS, is a system for automatic target detection, identification and information processing for use in human activities surveillance. This system consists of multiple sensors, and with control of their deployment and autonomous servo. Finding the optimal configuration for these sensors in order to capture the target objects and their environment to a required specification is a crucial problem. With a stereo pair of sensors, the 3D space can be discretized by an iso-disparity surface, and the depth reconstruction accuracy of the space is closely related to the iso-disparity curve positions. This paper presents a method to enable planning the position of these multiple stereo sensors in indoor environments. The proposed method is a mathematical geometry model, used to analyze the iso-disparity surface. We will show that the distribution of the iso-disparity surface and the depth reconstruction accuracy are controllable by the parameters of such model. This model can be used to dynamically adjust the positions, poses and baseline lengths of multiple stereo pairs of cameras in 3D space in order to get sufficient visibility and accuracy for surveillance tracking and 3D reconstruction. We implement the model and present uncertainty maps of depth reconstruction calculated while varying the baseline length, focal length, stereo convergence angle and sensor pixel length. The results of these experiments show how the depth reconstruction uncertainty depends on stereo pair's baseline length, zooming and sensor physical properties.

***Keywords**: Iso-disparity Surface, Multiple Surveillance Sensors, Stereo Vision, Sensor Configuration, 3D Reconstruction.*

1. Introduction

The human ability to process visual information may be extended with the help of advanced technologies. The Intelligent Vision Agent System, IVAS, is one such high-performance autonomous distributed vision and information processing system. The system involves collecting data with different levels of speed and accuracy, in order to reconstruct 3D information for security, health care, and surveillance applications. The system is able to focus on the important and informative parts of a visual scene by dynamically controlling the pan-tilt-zoom of a stereo pair. For such a system, the critical problem is to find the optimal configurations of sensors and to gain the required level of reconstruction accuracy. The stereo pair camera's profile, such as baseline length, convergence angle, pixel size, and focal lengths, are the most influential factors in determining the accuracy of 3D reconstruction. The effect of these parameters on the

3D reconstruction can be analyzed from the shapes and positions of the iso-disparity surfaces in 3D space.

The shape of the iso-disparity surfaces for general stereo configurations was first studied by Pollefeys et al., [1]. They described how the iso-disparity surfaces characterize the uncertainty and discretization in stereo reconstruction. A qualitative analysis of iso-disparity curves is also given. Although the geometry of these surfaces is well known in the standard stereo case, i.e. the front-parallel camera, there is a lack of analysis for the general stereo configuration. Also, the quantitative analysis of the iso-disparity surfaces has not often been studied.

This paper presents the iso-disparity surface geometry model, which is important in optimizing the stereo pair's configuration and in precisely anticipating the depth reconstruction accuracy. In addition, the model can also be used to make assumptions for many stereo algorithms, since these algorithms make hypothesis relying on the disparity range, i.e. the matching algorithm, [2]. In active vision, by estimating the disparity, control of the stereo convergence angle has already been introduced, [3]. The iso-disparity geometry model in active vision can help to select the disparity range according to surface geometry of the target.

Consideration of the iso-disparity when calculating the reconstruction uncertainty has been discussed by Völpel and Theimer, [4], where the disparity is considered in the *x*- and *y*- direction without using the epipolar geometry approach. In this paper, the disparity is defined along the epipolar line and the reconstruction uncertainty problem is solved by the iso-disparity geometry equations. This can be applied in more general cases.

A simple factor which helps to control the depth reconstruction accuracy is introduced in [5]. This paper improves the analysis of depth reconstruction accuracy into the stereo pair's *F*ield of *V*iew, FoV. This paper also considers the adjustment of stereo baseline for one stereo pair with a view to improving the depth reconstruction accuracy. The dynamic adjustment of stereo baseline for a parallel stereo pair was introduced in [6].

2. Definitions and problem formulation

The depth reconstruction accuracy can be controlled by adjusting the intervals of iso-disparity surface. This gives the possibility of planning a multiple sensor system, which can be implemented to observe human activities in 3D space with the required depth accuracy.

2.1 Definitions

Similar to the human eyes, stereo vision observes the world from different points of view. Two images are needed which are the fused to obtain a depth perception of the world. Any point in the world scene is captured in these two images as corresponding points which lie on the corresponding epipolar lines. It is necessary to define three terms related to the depth reconstruction: disparity, depth reconstruction uncertainty and depth reconstruction accuracy.

Disparity in this paper refers to the displacement of corresponding points along the corresponding epipolar lines for a common scene point, [1]. In the case where epipolar

lines are horizontal the disparity is measured directly from the difference of the corresponding points' coordinates. The inverse projection of all possible image points with the same disparity will reconstruct the iso-disparity surfaces in 3D space.

Depth reconstruction uncertainty is defined as the intervals between discrete iso-disparity surfaces due to the discrete sensor. The depth reconstruction accuracy is the inverse of the depth reconstruction uncertainty.

2.2 Problem statement and main contributions

The depth reconstruction may be calculated from a stereo pair with an accuracy determined by the system configuration. The system configuration is defined by sensor resolution (pixel size), focal lengths, baseline length and convergence angle. To get a more accurate depth reconstruction, the stereo configuration can be adjusted within its limits. The proposed models are limited to a general parallel stereo pair with zooming and the convergence stereo pair.

The main contributions of the paper can be summarized as follows:

- To model the iso-disparity surfaces for a general parallel stereo pair with zooming, as function of the baseline length, focal lengths and sensor pixel size.
- To model the iso-disparity surfaces for a general convergence stereo pair as function of the baseline length, convergence angle, focal length and sensor pixel size.
- Using iso-disparity geometry surfaces to quantitatively analyze the depth reconstruction accuracy.

3. Problem analysis

The iso-disparity surfaces of a stereo pair may be simulated using synthetic methods. However such simulation is time consuming, and for planning real-time multi sensor system an easy mathematical model of the iso-disparity surfaces is needed.

There are two configurations for a stereo pair in common use. The first one is a parallel stereo pair in which the optical axes of the cameras are parallel. The cameras may have the same focal lengths or their focal lengths may be different, e.g. to get better reconstruction accuracy of a target placed at the boundaries of cameras' field of view. The second common configuration is a convergence stereo pair, where the optical axes cross at a fixation point. The simple mathematical models of iso-disparity surfaces for these configurations are analyzed in this chapter.

3.1 The iso-disparity surface of a parallel stereo pair

From the geometry of a parallel stereo pair, two cameras with parallel optical axes, with different focal lengths for left and right camera, f_L and f_R respectively, the iso-disparity plane for disparity $n\Delta D$ can be defined as:

$$z(x,n) = \frac{f_L - f_R}{n\Delta D}x + \frac{B}{2n\Delta D}(f_L + f_R), \tag{1}$$

where B is baseline length, n is integer number, ΔD is the disparity resolution. The planes are shown as the thin green lines in Fig. 1(a) and Fig. 1(c).

All the iso-disparity planes intersect with the *xy*-plane (the stereo pair baseline is a part of *x*-axis), and converge to the straight line:

$$x = \frac{B}{2}\frac{f_L + f_R}{f_R - f_L}, \quad z = 0 \qquad (f_L \neq f_R). \tag{2}$$

It is clear from equation (1) that when the focal lengths are equal $f_L=f_R=f$, z is independent of x and the iso-disparity planes are parallel to *xy*-plane, see the thin green lines in Fig. 1(b).

From the inverse projection of image points and the triangulation method, using the Epipolar Geometry Toolbox, [7], we can get the synthetic iso-disparity surfaces. Fig. 1 shows the synthetic disparity surfaces (the bold red lines) and the plots from equation (1) (the thin green lines). Here the baseline length B is 30 cm and the disparity resolution ΔD=0.04 cm, or ten sensor pixel lengths where p=0.004 cm. Fig. 1(a) and Fig. 1(c) are plotted for the parallel stereo pair with different focal lengths. The parallel iso-disparity planes for parallel stereo pairs with same focal lengths are shown in Fig. 1(b). The synthetic simulation and calculating the equation give the same results.

3.2 The iso-disparity surface of convergence stereo pair

Let us consider two cameras with a convergence angle α_c, where $\alpha_{cL0}=\alpha_{cR0}=\alpha_c$ for the left and right camera respectively, with the angles rotated inwards to achieve a fixation point FP_0, as in Fig. 2. If the point TP_0 lies on the baseline's axis of symmetry, then the angles, (ψ_{L0}, ψ_{R0}), are the angles between the visual lines and a line perpendicular to the baseline. The zero disparity circle is defined by the fixation point and the left and right camera position points C_L and C_R. This circle is known as Vieth-Müller circle, and is a projection of the horopter, [8].

The iso-disparity surface is a cylinder whose cross section on the *xz*-plane is a conic that passes through both the centers of projection C_L, and C_R, and the point M_∞. M_∞ is a point imagined at infinity in both images, which can be obtained from the intersection of the normals to the optical axes, going through the projection centrals, [1]. It is possible to prove that for the case when α_{cL0},=$\alpha_{cR0}=\alpha_c$, the conic is an ellipse. We need to define the ellipse's five degrees of freedom. Three of these are determined by the points C_L, C_R and M_∞. One of the two remaining degrees is related to the point TP_0 with the disparity $n\Delta D$. The relationship between disparity $n\Delta D$ and focal lengths f_L and f_R for left and right cameras respectively, is the last required degree of freedom. If the disparity $n\Delta D$ and focal lengths f_L and f_R are known, the unique ellipse can be determined.

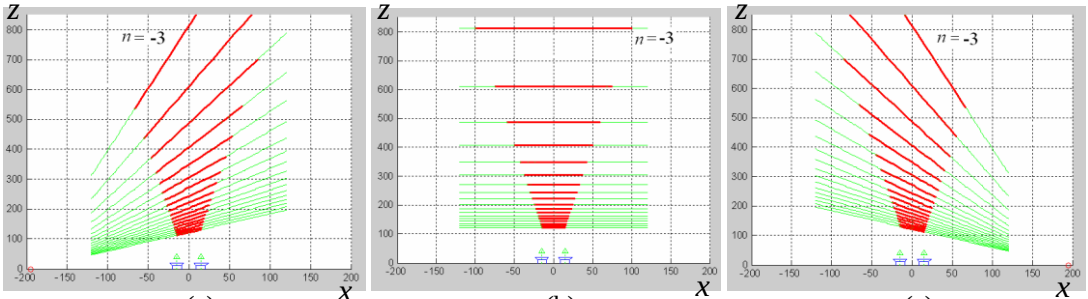


Fig. 1. Iso-disparity planes for parallel stereo pair from synthetic (bold red line) and a plot of the mathematical model from equation (1). The lines are plotted with steps of 10 pixels. (a) Cameras with different focal lengths, f_L = 3.5 cm, f_R = 3.0 cm for left and right camera respectively. The convergence point is (-195 cm, 0) on xz-plane. (b) Cameras with the same focal length of 3.25 cm. (c) Cameras with different focal lengths f_L = 3.0 cm, and f_R = 3.5 cm for left and right cameras respectively, the convergence point is (195 cm, 0) on xz-plane.

The iso-disparity surface of discrete disparity $n\Delta D$ for a convergence stereo pair (C_L, C_R) with the same focal length f and same convergence angles α_{cL0},=α_{cR0}=α_c, describes a cylinder, the ellipses being cross sections of this cylinder on the xz-plane with centres in $0_e(x_{0e}(n), z_{0e}(n))$:

$$\frac{(x - x_{0e}(n))^2}{a^2} + \frac{(z - z_{0e}(n))^2}{b^2} = 1. \qquad (3)$$

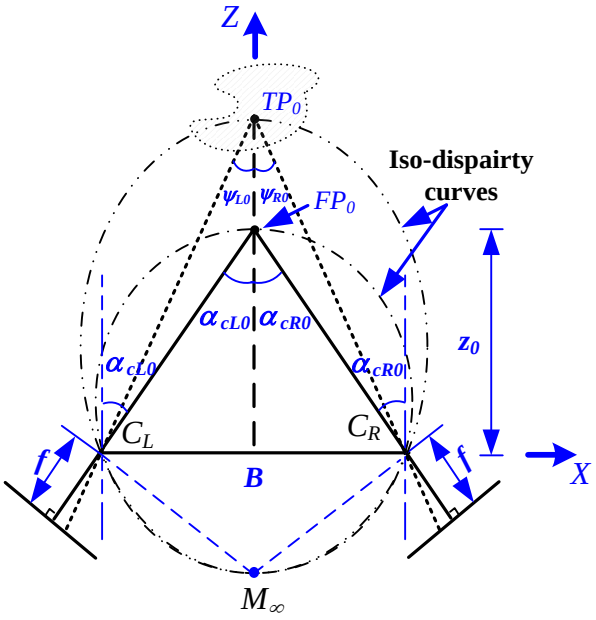


Fig. 2. An example of the iso-disparity curves for the convergence stereo pair in the plane defined by the cameras optical axes. z_0 is the distance from the fixation point to the baseline, f is the focal length.

For the chosen coordinates x_{0e}=0 and z_{0e}=b-Btanα_c/2:

$$\frac{x^2}{a^2}+\frac{\left(z-\left(b-\frac{B}{2}\tan\alpha_c\right)\right)^2}{b^2}=1, \tag{4}$$

where B is the baseline length and α_c is the stereo convergence angle. The ellipse half-axis along the z-axis $b(n,\Delta D,B,f,\alpha_c)$ depends on the discrete disparity $n\Delta D$, baseline length B, focus length f and convergence angle α_c and is described as:

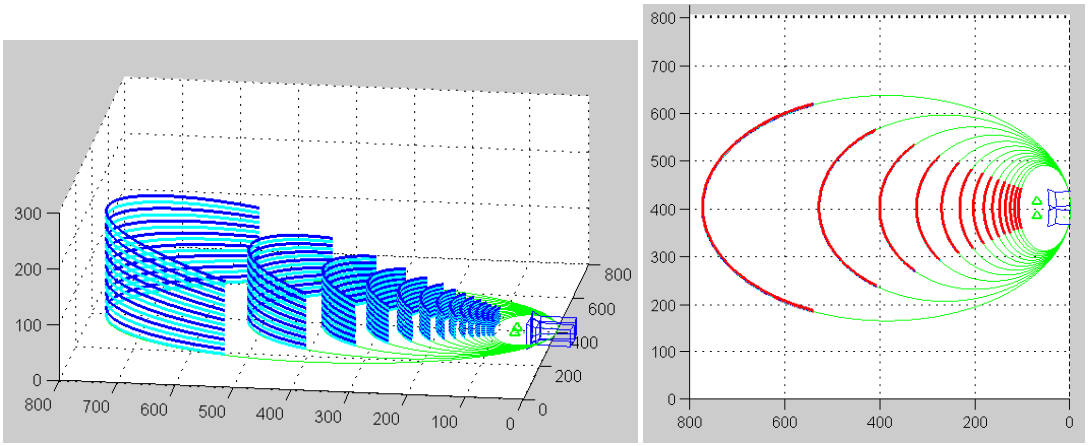
$$b=\frac{\frac{B}{2}}{\sin 2\alpha_c-\frac{n\Delta D}{f}\cos^2\alpha_c}\quad . \tag{5}$$

The ellipse half-axis along the x-axis $a(n,\Delta D,B,f,\alpha_c)$ can be found from the relationship:

$$\left(\frac{b}{a}\right)^2=\tan\alpha_c\,\frac{1+\frac{n\Delta D}{2f}\tan\alpha_c}{\tan\alpha_c-\frac{n\Delta D}{2f}}=\frac{\tan\alpha_c}{\tan\psi_c}, \tag{6}$$

where ψ_c = ψ_{L0} = ψ_{R0}.

The result of the synthetic stereo pair simulation with a baseline length 50 cm, focal lengths 2.5 cm and disparity resolution ΔD=0.04 cm is shown in Fig. 3. The synthetic iso-disparity surfaces (bold red lines) can be compared with ellipses from the equations (4)-(6) (thin green lines). Fig. 3 shows both the synthetic iso-disparity surfaces and the ellipses from the equation (4) in 3D space, in perspective view in Fig. 3(a) and top view



(a) (b)

Fig. 3. Simulation results of iso-disparity surfaces for a stereo pair from the synthetic model (bold red line) and mathematical model from equation (4) (green line) with convergence angle, α_c=4°, the baseline length B=40 cm, the focal lengths f=2.5 cm and disparity resolution ΔD=0.04 cm (a) perspective view, and (b) top view.

in Fig. 3(b). Both results match each other perfectly.

4. Approach

Since the gaps between iso-disparity surfaces represent the discretization uncertainty in 3D space, we can generate a 3D depth reconstruction uncertainty map of a particular stereo pair's configuration using the iso-disparity surface geometry equations (4)-(6). Also, it is possible to generate such a map in 2D on the optical axes plane. This map can be used for the optimization of the stereo setup configuration. Generation of the 2D uncertainty map for a stereo pair configuration can be done in the following three steps.

Firstly, the plane has to be covered by the stereo pair's FoV, [9]. The area is sampled using small grids covered by the stereo pair.

Secondly, an iso-disparity curve on the optical axes plane should be calculated, passing through each grid point. Knowing that the curve will have a canonical shape, five points are needed. Two of these points can be the grid point and its symmetrical point, with respect to the symmetry axis of the baseline. The three others points are C_L, C_R and M_∞. For a convergent stereo pair, the ellipse axes a and b can be found using the ellipse fitting algorithm, [10]. Then using equation (5), the two closest ellipses with discrete disparity values $n\Delta D$ and $(n+1)\Delta D$ respectively, can be found, where the disparity resolution ΔD is one sensor pixel length.

Finally, the depth reconstruction uncertainty can be calculated as the interval between the iso-disparity surfaces, with the disparity values, $n\Delta D$ and $(n+1)\Delta D$ as the distance between the intersections of these two iso-disparity surfaces, and the line through the grid point and M_∞.

5. Results

The simulations presented were performed in MATLAB 7.0, and cover a rectangular area of (800 cm × 800 cm). This case study illustrates how depth reconstruction uncertainty in stereo coverage varies with the target distance for a given stereo baseline length, focal length, and sensor pixel length. The results are presented in Fig. 4, where the cameras optical axes are in the xz-plane. The depth reconstruction uncertainty is specified by the positive y-axis of the coordinate. However, this uncertainty analysis shows only the area covered by the stereo pair's FoV. To scale the uncertainty on the optical axes plane, a colour map is used. The lowest uncertainty is indicated by the blue colour and the highest uncertainty by the red colour. In order to increase the readability of the iso-disparity curves, the contour is plotted with ten pixel lengths disparity resolution. The map of the iso-disparity curves is generated with baseline length 40 cm, focal length 3.5 cm and pixel length p=0.004 cm, stereo convergence angle, α_c=4° and the FoV is approximately 54°. This case study proves that the depth reconstruction uncertainty increases as the distance to the target increases.

To show the discrete properties of depth reconstruction uncertainty, the map of the iso-disparity curves with suitable baseline length and pixel length is shown in the Fig. 5. The figure shows only half of FoV, with a cross section along the ellipses' axes perpendicular to the baseline. The discretization step increases with the target distance.

An exact illustration as to how the depth reconstruction uncertainty varies with the baseline lengths, focal lengths, sensor pixel length and stereo convergence angle, is shown in Fig. 6 and Fig. 7. Fig. 6(a) shows that the relative depth reconstruction uncertainty, relative to the target distance, decreases when the baseline length increases. The relative uncertainty decreases slowly for a baseline above about 40 cm. Its minimum value tends to be constantly between 0.5% and 1.5% for target distances of 200 cm and 800 cm respectively. At the same time, for a baseline of about 10 cm, the uncertainty varies between 10% and 2.5% for the respective target distances.

The change of the relative uncertainty versus the focal length is similar to that of the baseline length; see Fig. 6(b). For a focal length of longer than 3.5 cm the increase of the uncertainty is relatively slow. Its minimum tends to be constantly between 1.5% and 0.4% for target distances of 200 cm and 800 cm respectively. Meanwhile, for a focal length of 1 cm, the uncertainty varies between about 9% and 2% for the respective target distances.

Furthermore, Fig. 7(a) illustrates the linearly relation of the relative uncertainty and the sensor pixel length. Within the range from 0.001 cm to 0.006 cm, the relative uncertainty varies from 0.2% to 3.5% and depends also on the target distance.

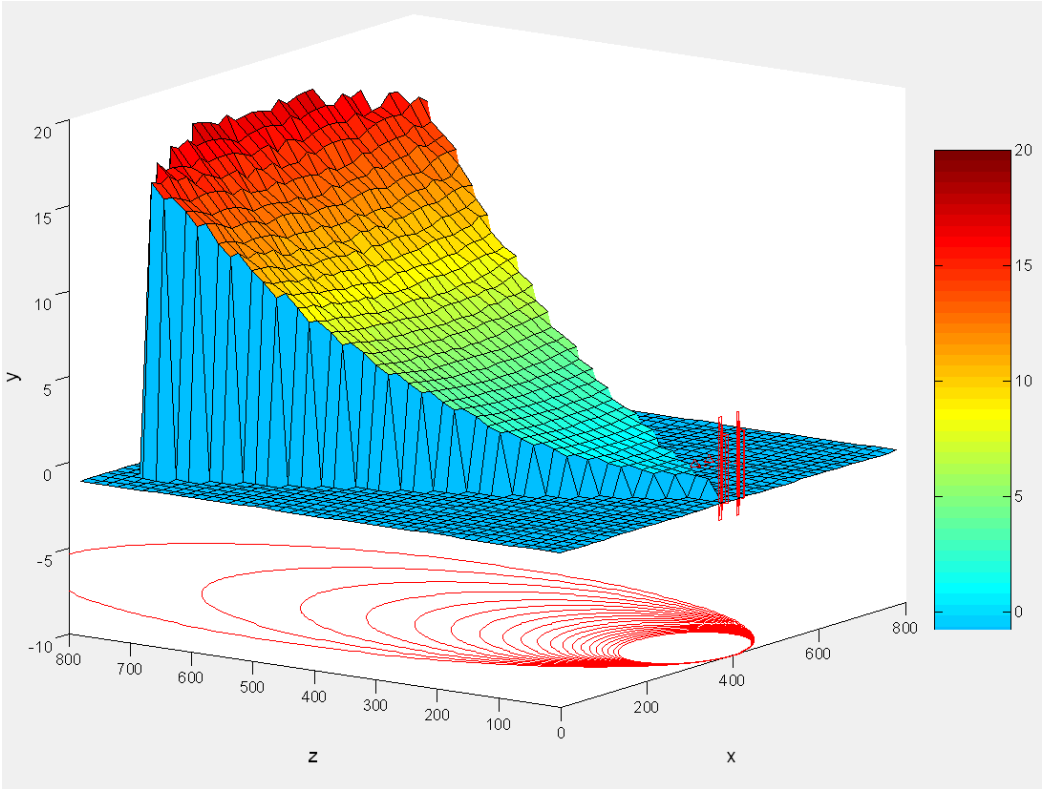


Fig. 4. The depth reconstruction uncertainty map for a stereo pair's FoV, where B=40 cm, f=3.5 cm, p=0.004 cm.

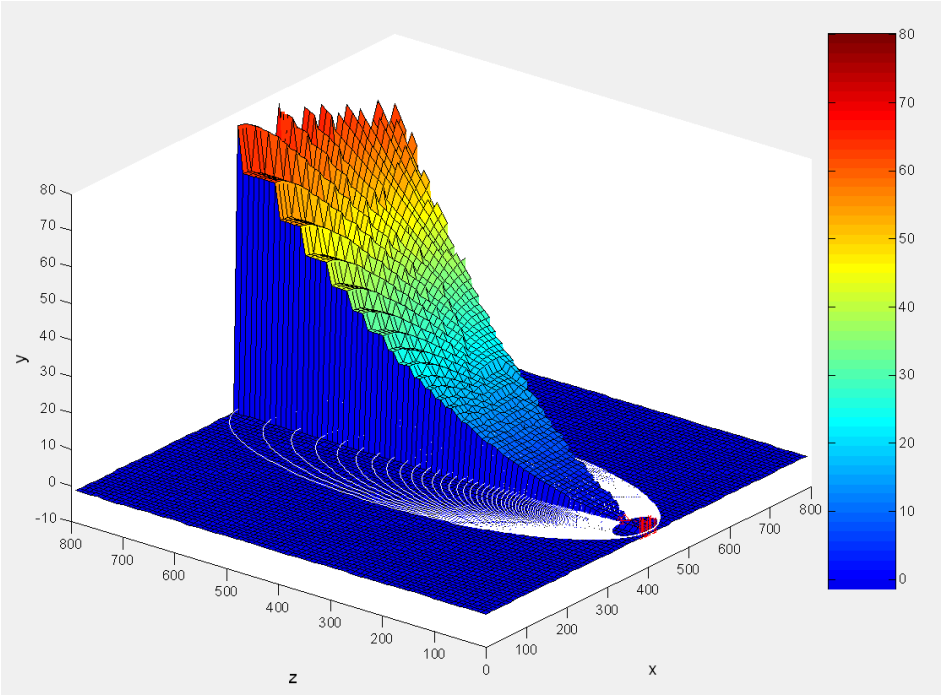
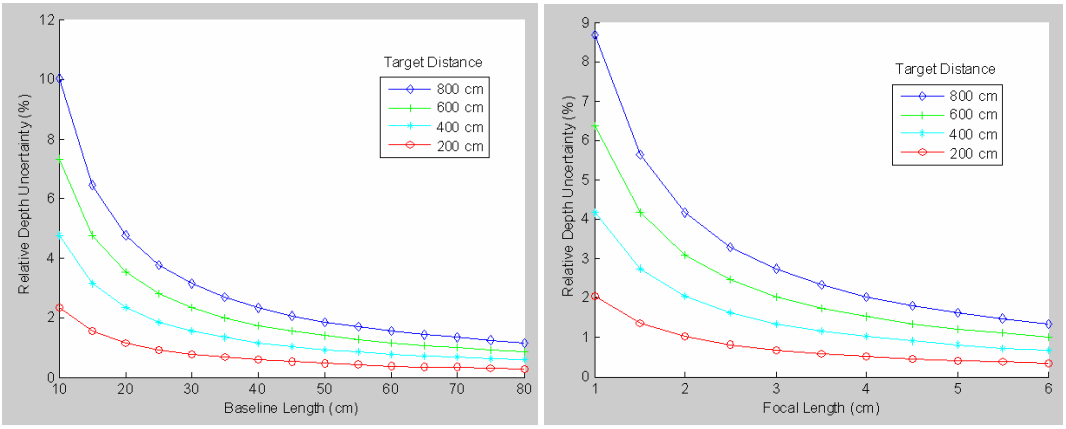


Fig. 5. The depth reconstruction uncertainty map for a stereo pair's half FoV, where B=20 cm, f=3.5 cm, p=0.008 cm.

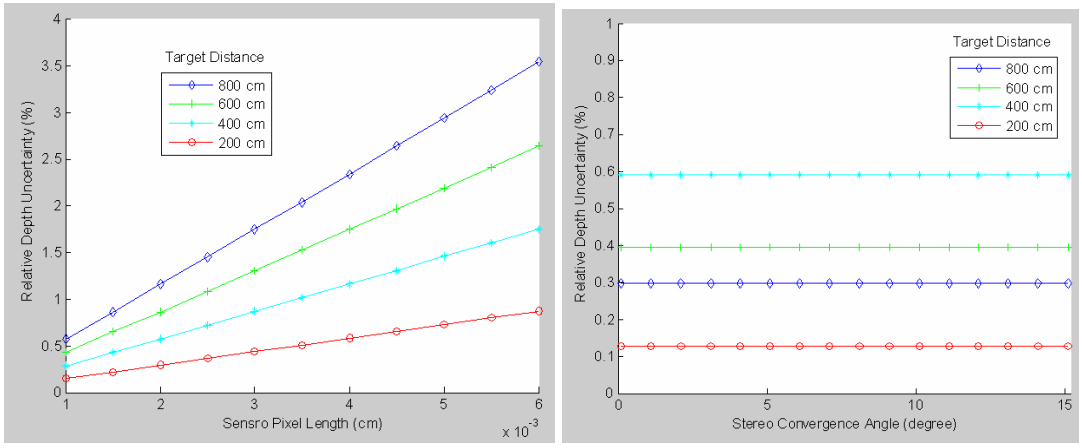
Fig. 7(b) shows that the stereo convergence angle has a slight influence on the uncertainty but this also depends on the target distance.

Fig. 8(a) and Fig. 8(b) illustrate the variation of the uncertainty when both the focal length and the baseline length are changed for two different target distances, 200 cm and 600 cm, respectively. The uncertainty increases significantly when the baseline length decreases below 40 cm, independent of the location of the target within the FoV.



(a) (b)

Fig. 6. The uncertainty varies with the baseline length, focal length, sensor pixel length, and stereo convergence angle. The distance from the target to the camera is 800 cm, 600 cm, 400 cm and 200 cm, respectively. They are marked by different type of lines. The uncertainty varies with (a) the baseline length; (b) the focal length.



(a) (b)

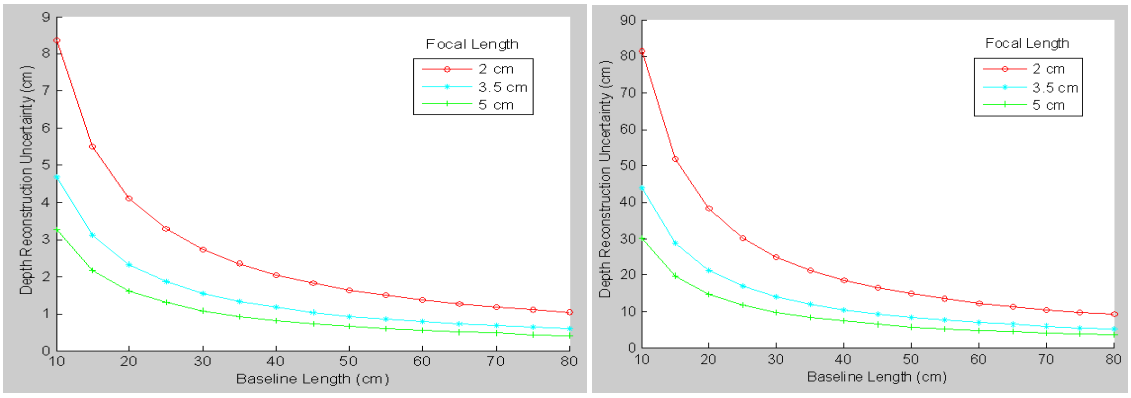
Fig. 7. The uncertainty varies with the baseline length, focal length, sensor pixel length, and stereo convergence angle. The distance from the target to the camera is 800 cm, 600 cm, 400 cm and 200 cm, respectively. They are marked by different type of lines. The uncertainty varies with (a) the sensor pixel sizes; (b) the convergence angle.

Also, a significant increase in the uncertainty is visible for a focal length below 3.5 cm.

The relative accuracy is similar for a target located in different positions, but its absolute value is more significant for a target further from the stereo pair. In order to fulfil the reconstruction accuracy requirement for a faraway target, the focal length or baseline has to be adjusted. A longer focal length can be used to compensate for a shorter baseline. And also, in general, the longer the baseline is, the more difficult the matching becomes.

6. Conclusion

The planning and control of multi stereo pairs' baselines, positions and poses for



(a) (b)

Fig. 8. The uncertainty varies with both the focal length and the baseline length. The focal lengths are 2 cm, 3.5 cm and 5 cm, respectively, marked by different type of lines. The target is (a) 200 cm; (b) 600 cm faraway from the camera.

surveillance and tracking purposes, e.g. in supermarkets, museums, the home environment, and especially in situations which require stereo data to reconstruct 3D with a required accuracy, are possible fields of application. The proposed approach may be used in the dynamic control of stereo pair's baseline, and cameras' corresponding positions and poses, to observe a moving target.

The analysis presented shows that the depth reconstruction accuracy varies more significantly with respect to the target distance to baseline, baseline length and focal length than to the convergence angle. Small changes in stereo convergence angle do not affect the depth accuracy overly much, especially when the target is placed centrally. On the other hand it can have a great impact on the shape of the iso-disparity curves. From the proposed iso-disparity mathematical model we can get reliable control of the iso-disparity curves' shapes and intervals by using the systems configuration and target properties.

To achieve a more accurate 3D reconstruction of the target, it is better to bring the target to an area with a small depth reconstruction uncertainty. Furthermore, the controllable disparity distribution can determine and verify the assumptions which are used in stereo algorithms.

Future work could focus on the dynamical adjustment of the configuration of a stereo pair according to the target shape and position. The iso-disparity geometry model could also be used as a guide for stereo rectification or matching.

References

[1] M. Pollefeys and S. Sinha, "Iso-disparity Surfaces for General Stereo Configurations", in: Proc. of the 6th European Conf. on Computer Vision, 2004.

[2] M. Christian, S. Robert, "Adaptive Area-based Stereo Matching", in: Proc. of SPIE, Three-Dimensional Image Capture and Applications, vol. 3313, pp. 14-24, 1998.

[3] H. J. Kim, M. H. Yoo, and S.W. Lee, "Dynamic Vergence Using Disparity Flux", in: Proc. of the 1th IEEE Int. Workshop of the Biologically Motivated Computer Vision, 2000.

[4] B. Völpel and W. M. Theimer, "Localization Uncertainty in Area-Based Stereo Algorithms", IEEE Transactions on Systems, Man, and Cybernetics, vol. 25, no. 12, 1995.

[5] J. Chen, S. Khatibi and W. Kulesza, "Planning of A Multi Stereo Visual Sensor System- Depth Accuracy and Variable Baseline Approach", in: Proc. of IEEE Computer Society 3DTV-Con, the True Vision Capture, Transmission and Display of 3D Video, Kos, Greece, 2007.

[6] Y. Nakabo, T. Mukai, Y. Hattori, Y. Takeuchi and N. Ohnishi, "Variable Baseline Stereo Tracking Vision System Using High-Speed Linear Slider", in: Proc. of the IEEE Int. Conf. On Robotics and Automation, 2005.

[7] G. L. Mariottini, D. Prattichizzo, "The Epipolar Geometry Toolbox: Multiple View Geometry and Visual Servoing for Matlab", in: Proc. of IEEE Int. Conf. on Robotics and Automation, 2005.

[8] K. Ogle, Researches in Binocular Vision, W.B. Saunders Company, Philadelphia & London, 1950.

[9] J. Chen, S. Khatibi and W. Kulesza, "Planning of A Multi Stereo Visual Sensor System for A Human Activities Space", in: Proc. of the 2nd Int. Conf. on Computer Vision Theory and Applications, 2007.

[10] R. Halif and J. Flusser, "Numerically Stable Direct Least Squares Fitting of Ellipses", in: Proc. of the 6th Int. Conf. Computer Graphics and Visualization, pp. 125 – 132, 1998.

Paper II

Depth Reconstruction Uncertainty Analysis and Improvement –The Dithering Approach

Authors:

Jiandan Chen, Siamak Khatibi and Wlodek Kulesza

Reformatted version of paper originally published in:

Elsevier Journal of Image and Vision Computing, vol. 28, no. 9, pp. 1377-1385, September, 2010

Depth Reconstruction Uncertainty Analysis and Improvement –The Dithering Approach

Jiandan Chen, Siamak Khatibi and Wlodek Kulesza

Abstract

The depth spatial quantization uncertainty is one of the factors which influence the depth reconstruction accuracy caused by a discrete sensor. This paper discusses the quantization uncertainty distribution, introduces a mathematical model of the uncertainty interval range, and analyzes the movements of the sensors in an Intelligent Vision Agent System. Such a system makes use of multiple sensors which control the deployment and autonomous servo of the system. This paper proposes a dithering algorithm which reduces the depth reconstruction uncertainty. The algorithm assures high accuracy from a few images taken by low-resolution sensors. The dither signal is estimated and then generated through an analysis of the iso-disparity planes. The signal allows for control of the camera movement. The proposed approach is validated and compared with a direct triangulation method. The simulation results are reported in terms of depth reconstruction error statistics. The physical experiment shows that the dithering method reduces the depth reconstruction error.

Keywords*: Depth Reconstruction, Quantization Uncertainty, Dither, Iso-disparity Planes.*

1. Introduction

The human ability to process visual information may be extended with the help of some advanced technologies. The Intelligent Vision Agent System, IVAS, is one such high-performance autonomous distributed vision and information processing system, [1]. The system collects data in order to reconstruct 3D information for security, health care, medical and surveillance applications, and so on. It focuses on the important and informative parts of a visual scene by dynamically controlling the pan-tilt-zoom of a stereo pair. For such a system, the critical problem is to find the optimal configurations of sensors and to gain the required reconstruction accuracy. By adjusting the stereo pair's profile, such as baseline, convergence angle, focal lengths and pixel size, the depth reconstruction accuracy can be improved, [2], [3]. When determining the accuracy of 3D reconstruction, depth spatial quantization is one of the most influential factors. This type of factor cannot be reduced even in more accurate measurements. For instance, increasing the image resolution by reducing the sensor pixel size is of limited use since the signal-to-noise ratio, SNR, is then reduced, and also because of the restricted sensitivity of the sensor itself. The selection of an optimal sensor pixel size is discussed by Chen et al., [4].

How to reconstruct a super-resolution image from the low-resolution images has been the focus of much research in recent years. To overcome the digital camera sensor pixel size limitation, attempts have been made to combine the information from a set of slightly different low-resolution images of the same scene and use them to construct a higher-resolution image. The methods used can be categorized as belonging to frequency domain analysis, statistical analysis, and the geometrical interpolation approach, [5], [6]. Francisco and Bergholm introduced a method whereby the stereo vergence could be changed by shifting the camera sensor chip, [7]. Also, Ben-Ezra et al. proposed a jitter camera to minimize the motion blur in the reconstruction of video super-resolution, [8]. The quantization uncertainty of depth reconstruction for parallel and convergence stereo cameras is discussed by Chen et al., [9]. However, there has been little work on reducing the depth reconstruction uncertainty based on a combination of low-resolution images. In this paper, the dithering algorithm combines dithering and iso-disparity analyses to find the optimal number of low-resolution images used to reduce the depth reconstruction uncertainty.

Dithering is a well-known technique that is applied in analogue-to-digital converters, *ADCs*. To control the output error, a proper dither signal is added to the sequence before the quantizer. A mathematical model of dithered quantization was introduced by Wannamaker et al., [10], [11], Wagdy, [12], and Carbone and Petri, [13], and demonstrated that the system resolution decreased below the least significant bit. Although the *ADCs* analysis is performed in the time domain, our case requires that the depth reconstruction is analyzed, and the dither signal added, in the space domain.

For instance, many astronomical cameras use the dithering method, which in this case involves multiple exposures of the same field with small shifts between the exposures. This feature enables reconstruction of some of the information which has been lost because of the digital camera's discretization, [14]. In relation to this, Yang, [15], has addressed the problem that the quantization error becomes serious when the size of the pixel is significant compared to the allowable measurement tolerance. In his paper, the error analysis and sensor position optimization for the inspection of the edge line are explored. Liu and Ehrich, [16], also use dithering to locate a subpixel edge in a binary image. Klarquist and Bovik, [17], presented a vergent active stereo vision system to recover a high depth resolution by accumulating and integrating a multiresolution map of the surface depth over multiple successive fixations.

In general, pixels of camera sensor are uniformly distributed in a two dimensional array. The projection of each 3D point in the scene is approximated to the center of the nearest pixel; the resulting error is referred to as a quantization error. In stereo, the quantization error generates an uncertainty in the depth estimation at each 3D point. Basu and Shabi, [18], have introduced a model using stereo cameras with a non-uniform resolution sensor based on an optimal estimation of the 3D points' locations. Furthermore, Kil et al., [19], have used a laser scanner to reconstruct a high resolution 3D image of the target surface using hundreds of lower resolution scans as inputs. The lower resolution scans are randomly shifted, so that each of them has contributed information to the final model. The limitation of this approach is that a huge number of scans are required as inputs, and the improvement in accuracy cannot be controlled. Unlike Kil et al.'s approach, the algorithm proposed in this paper can control the depth reconstruction uncertainty. Farid, [20], introduced a method whereby the depth can be determined by placing a variable opacity optical attenuation mask directly in front of a

camera lens. The small continuum of viewpoints' change can be achieved by the choice of non uniformed masks as opposed to the discrete views obtained by shifting the cameras. The fast computation and the avoidance of correspondence points matching calculation that characterize the proposed approach can be amenable to a real-time implementation of the proposed algorithm based on shifting cameras in future work.

2. Problem statement and main contributions

2.1 Problem statement

At least two images are needed to obtain a depth map of the world, the accuracy of which is limited by the camera sensor resolution. When a digital camera takes an image, the scene perspective is projected onto a sensor plane. The sensor elements are arranged into two-dimensional arrays to represent the scene. The coordinates of the image are discrete and the resolution is in pixels. This leads to a depth reconstruction quantization uncertainty in the representation of the spatial position of a scene. The depth reconstruction uncertainty depends on the pixel size, [9], and becomes a serious issue when the pixel size is large compared to the allowable tolerance in the depth reconstruction. Therefore, it becomes crucial to find a method that can help reduce the depth reconstruction uncertainty. The proposed algorithm overcomes the sensor resolution limitation and enables more accurate depth reconstruction based on the dithering method. This makes it possible to plan a multiple sensor system which can be implemented and help observe a target object in 3D space with the required depth accuracy.

2.2 Main contributions

This paper contributes to current research by:

- Analyzing the depth reconstruction uncertainty distribution;
- Providing a method and its mathematical model that reduces the depth uncertainty;
- Estimating and generating discrete dither signals by an analysis of the iso-disparity planes;
- Identifying an accepted 3D reconstruction space with defined accuracy;
- Implementing a simple and robust dithering algorithm that significantly reduces the depth reconstruction quantization uncertainty.
- Identifying an accepted 3D reconstruction space with defined accuracy;

3. Problem analysis

Depth reconstruction may be calculated from stereo pair cameras with an accuracy determined by the system configuration, which is defined by the sensor pixel resolution Δ (a square pixel of the size $\Delta \times \Delta$), the focal length f, and the baseline length, B, for a general parallel stereo pair. To get a more accurate depth reconstruction, the stereo configuration can be adjusted within its limits. Since the quantization uncertainty is caused by the properties of a digital sensor, a dithering method is a suitable method to improve the depth reconstruction accuracy.

3.1 The stereo pair geometric model and the image quantization uncertainty

Let us consider two cameras with parallel optical axes. The pinhole camera model is used. The 3D cameras' space coordinates are shown in Fig.1 and their origin is located in the middle of the two cameras. The positive *Z*-axis has the same direction and is parallel with the cameras' optical axes. The distance between the camera centers C_l and C_r is the baseline length B and the cameras have the same focal length f. For a given 3D scene point, its projections onto the left and right image planes are (x_l, y_l) and (x_r, y_r) respectively. However, the projection points are approximated by the pixel centers on the left and right image, and denoted (x_{Ql}, y_{Ql}) and (x_{Qr}, y_{Qr}) for the left and the right camera respectively. The difference between the approximate and the exact projections is the image quantization uncertainty. The image quantization uncertainties for the left and right images are (q_{hl}, q_{vl}) and (q_{hr}, q_{vr}), which can be denoted as (q_h, q_v) in the horizontal and vertical directions respectively.

The quantized depth of a target point, Z_{nt}, is the distance from the baseline to the target point computed from the image matrix. From the geometry of a parallel stereo pair with the same focal length f, the quantized depth can be obtained from:

$$Z_{nt} = \frac{Bf}{|x_{Ql} - x_{Qr}|\Delta} = \frac{Bf}{n_t \Delta}, \tag{1}$$

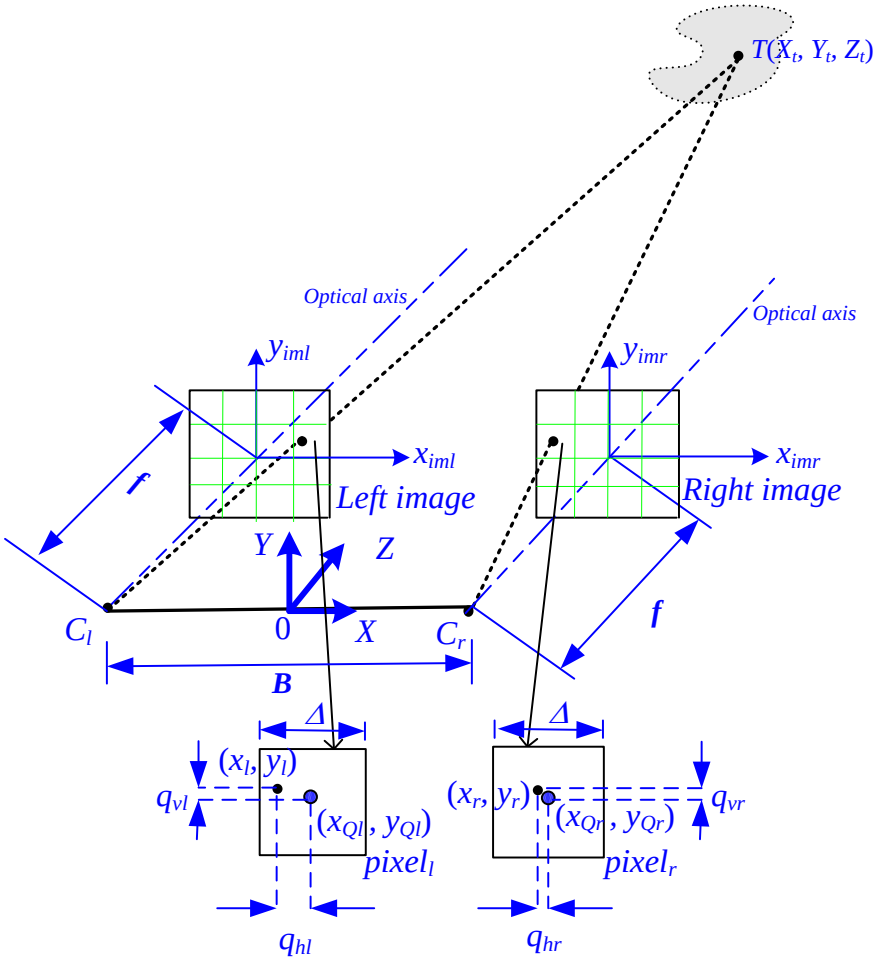


Fig. 1. Parallel stereo pair geometry and the image quantization uncertainty model.

where B is the baseline length, n_t is a target point disparity which is an integer number, calculated from $n_t = \left| x_{Ql} - x_{Qr} \right|$, and Δ is a pixel size. For a parallel stereo pair, the disparity n_t is a measure of the displacement of the corresponding projection points along image coordinates on the x-axis for a common scene point. It should be noted that the depth reconstruction uncertainty is directly related to the horizontal image coordinate. In this paper, we only focus on the one dimension uncertainty analysis which is along the horizontal image coordinate. From the perspective projection of a target point, T (X_t,Y_t,Z_t) onto the left and right images, the positions of the image coordinates on the x_{iml} and x_{imr} axes can be obtained from:

$$x_l = \frac{f}{Z_t}\left(X_t + \frac{B}{2}\right) \qquad x_r = \frac{f}{Z_t}\left(X_t - \frac{B}{2}\right). \tag{2}$$

These two exact projections x_l and x_r are approximated as x_{Ql} and x_{Qr} since the camera sensors can be considered as mid-tread quantizers. Since further analysis focuses on general quantization formulae (2), x_l and x_r are denoted as x_c. In general, the image quantization uncertainty consists of two components, one in the horizontal and one in the vertical direction. The uncertainties in the vertical direction for the rectified left and right images are the same. In our approach, we are interested only in the horizontal component, which we name the image horizontal quantization uncertainty, q_h, and describe as:

$$q_h(x_c) = \frac{\Delta}{2} - \Delta\left\langle \frac{x_c}{\Delta} + 0.5 \right\rangle, \tag{3}$$

where the symbol $\langle\cdot\rangle$ denotes the fractional part. The sawtooth-shaped function (3) is a periodical function of x_c with the period, Δ, and can be expressed as a Fourier series, [21]:

$$q_h(x_c) = \sum_{k=1}^{\infty} \frac{\Delta}{\pi k}(-1)^k \sin\left(2\pi k \frac{x_c}{\Delta}\right) \quad . \tag{4}$$

3.2 Disparity and the depth reconstruction quantization uncertainty

The point in space is projected onto the point on the image plane. The image horizontal coordinate of this point is quantized with the uniform image horizontal quantization uncertainty, q_h, of the range $\pm\Delta/2$. Assuming that the quantization uncertainties of the left and right images are uncorrelated, [22], the probability density functions, PDFs, of the image horizontal quantization uncertainties, p_{hl} and p_{hr}, for the left and right images respectively are:

$$\begin{cases} p_{hl}(q_{hl}) = \dfrac{1}{\Delta} & \text{where } -\dfrac{\Delta}{2} \le q_{hl} \le \dfrac{\Delta}{2} \\ p_{hr}(q_{hr}) = \dfrac{1}{\Delta} & \text{where } -\dfrac{\Delta}{2} \le q_{hr} \le \dfrac{\Delta}{2} \end{cases}, \tag{5}$$

where q_{hl} and q_{hr} are the left and right image horizontal quantization uncertainty respectively.

The disparity quantization uncertainty is $q_d = q_{hl} - q_{hr}$, and its probability density function can be described by the convolution of the probability density functions of the two horizontal quantization uncertainty PDFs:

$$p_d(q_d) = p_{hl}(q_{hl}) \otimes p_{hr}(q_{hr}) = \begin{cases} \frac{1}{\Delta^2}(q_d + \Delta), & -\Delta \le q_d < 0 \\ -\frac{1}{\Delta^2}(q_d - \Delta), & 0 \le q_d \le \Delta \\ 0, & \text{elsewhere} \end{cases}, \tag{6}$$

where $\otimes$ denotes convolution. The result of the convolution of two rectangular distributions is a triangular one. The quantization uncertainty interval of the disparity, q_d, is 2Δ and varies between $-\Delta$ and $+\Delta$.

For a parallel stereo pair, the displacement between two corresponding projection points on the left and right image planes along the image coordinates on the x-axis for a common scene point is called disparity and is depicted in Eq. (1) as the integer number n_t. For each disparity n_t, there is a corresponding iso-disparity plane in the 3D space for a stereo pair with a specified baseline length B, focal length f and pixel size Δ [9]. By using Eq. (1), we can estimate the target point's discrete depth, Z_{nt}, from the disparity n_t, which also corresponds to the n_t-th iso-disparity plane. The interval between the iso-disparity planes represents the depth reconstruction quantization uncertainty, q_{Zt} as the non-linear function of n_t, and can be denoted as:

$$q_{Z_t} = \frac{Bf}{n_t(n_t + 1)\Delta}. \tag{7}$$

3.3 The depth reconstruction uncertainty with dithering

In our proposed model, the left and right cameras are the quantizers. The dither signals add *noise* to the signals x_l and x_r prior to its quantization in order to change the statistical properties of the quantization, [11]. In our case, the signals x_l and x_r are the target point projections on the image planes along the X-axis. There are two possibilities for adding a dither signal to change the x_r and x_l signals: one is to shift the target features parallel with the X-axis. An alternative is to shift the camera sensor along the X-axis, which means that the quantization levels of the quantizer are changed. The proposed method is based on the movement of the camera sensor position along the X-axis.

The dither signal, d, is a discrete one and can be used to control the left and right cameras' position movement along the X-axis. This allows for a change in the projection of the target point in the range of $[-\Delta/2, \Delta/2]$ on the image coordinates x_{iml} and x_{imr}. The PDF of the discrete dither signal consists of the equiprobable impulses which number is greater than one. In this paper, we propose a two-stage discrete binary dither signal for each camera, which means four images to calculate a depth of the target feature with a reduced quantization uncertainty. In order to simplify the analysis, we assume that the mean value of the dither signals equals zero. The PDF of the discrete dither signal, D, is:

$$f(D) = \frac{1}{2}\left(\delta\left(D - \frac{D_a}{2}\right) + \delta\left(D + \frac{D_a}{2}\right)\right), \tag{8}$$

where $\delta(\cdot)$ represents the Dirac delta function; D_a is the discrete binary dither signal amplitude and equal to the displacement of the target point projection on the image plane.

The mean value of the image horizontal quantization uncertainty, $\bar{q}_h$, with the discrete binary dither signal D is:

$$\bar{q}_h(x_c) = E(q_h(x_c + D)) = \int_{-\infty}^{\infty} q_h(x_c + D) f(D) \mathrm{d}D, \tag{9}$$

where $E(\cdot)$ is the expected value operator. By using Eq. (4), we can get [21]:

$$\begin{aligned} \bar{q}_h(q_h, D_a) &= \sum_{k=1}^{\infty} \frac{\Delta(-1)^{k+1}}{\pi k} \cos\left(\pi \frac{k}{\Delta} D_a\right) \cdot \sin\left(2\pi \frac{k}{\Delta} q_h\right) \\ &= \sum_{k=1}^{\infty} \frac{\Delta(-1)^{k+1}}{2\pi k} \left[\sin \pi \frac{k}{\Delta}(2q_h - D_a) + \sin \pi \frac{k}{\Delta}(2q_h + D_a)\right], \end{aligned} \tag{10}$$

where x_c can be substituted by $-q_h$ due to its periodicity with the period, Δ, and q_h varies between -0.5Δ and 0.5Δ. When Δ is normalized into 1 pixel, Fig. 2 shows the plot of the function, $\bar{q}_h$, versus q_h with a range from -0.5Δ to 0.5Δ for five different amplitude values of the discrete dither signal, D_a=[0.1Δ, 0.4Δ, 0.5Δ, 0.6Δ, 0.9Δ].

From Fig. 2, it is clear that the mean image horizontal quantization uncertainty interval of an image with the discrete binary dither signal amplitude $D_a = 0.5\Delta$ has a minimum interval of 0.5Δ and varies in the range of -0.25Δ to 0.25Δ. Applying the result to Eqs. (5) and (6), the interval of the mean disparity quantization uncertainty, $\bar{q}_d$, becomes Δ and it varies in the range of [-0.5Δ, 0.5Δ]. Since we know from Eq. (6) that the interval of the disparity quantization uncertainty, q_d, without dithering is 2Δ in the range of [$-\Delta$, Δ], the interval of the disparity quantization uncertainty is reduced by half by applying the dithering method.

3.4 The dither signal generation

The dither signal is used to move the camera and control the displacement of a 3D

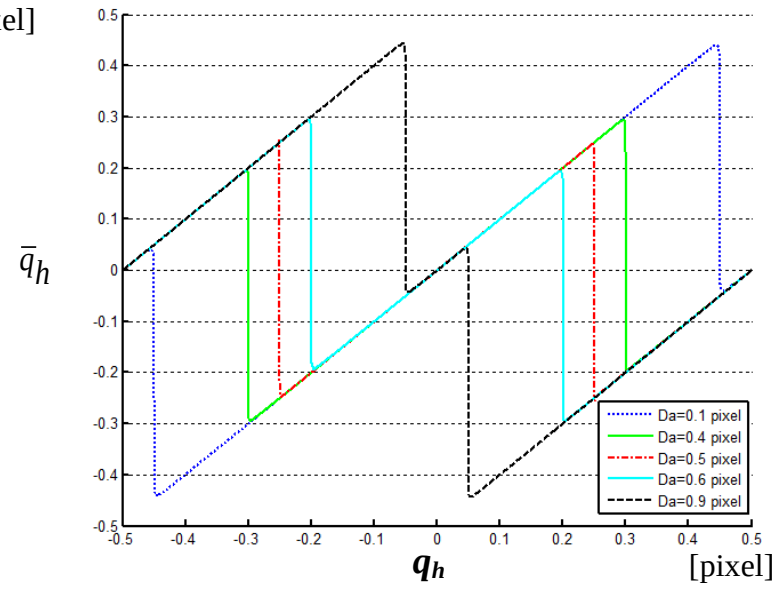


Fig. 2. The mean image horizontal quantization uncertainty with the dither signal, $\bar{q}_h$, versus the image horizontal quantization uncertainty q_h for different amplitude values of the dither signal amplitude D_a.

point projection on the image coordinate. To shift a point projection on the image coordinate, the camera displacement varies according to the target point's distance to the camera's baseline B and depends on the camera parameters, such as focal length f and pixel size Δ.

In the 3D space, the depth reconstruction uncertainty is represented by the intervals between the iso-disparity planes according to Eq. (7). The uncertainty can be reduced by half when one places the new iso-disparity planes into the middle of the old disparity planes. To get new iso-disparity planes, we can increase or decrease a baseline, which can be accomplished by a single camera movement. The required displacement of the target point projection on the image coordinate, which is also equaled to the dither signal, $D_a(n_t)$, can be calculated from:

$$D_a(n_t) = \frac{n_t}{2(n_t+1)}\Delta. \tag{11}$$

The target displacement can be yielded by increasing or decreasing of the baseline according to:

$$\Delta B_t = \frac{B}{2(n_t+1)} = \frac{B \cdot D_a(n_t)}{n_t \cdot \Delta}. \tag{12}$$

In Fig. 3a, red solid lines show the parallel iso-disparity planes for n from 30 to 200. Here, the stereo pair baseline length B is 50 mm, the pixel size Δ = 12.9 µm and both cameras have the same focal length f = 25 mm. For instance, if the target point's disparity n_t = 49, the measured depth is Z_{49} = 1997 mm. To shift the iso-disparity plane further away from the baseline by a half of the depth reconstruction quantization uncertainty, $q_{Zt}/2$, the baseline length has to increase by 0.51 mm. In Fig 3, green dashed lines show the new iso-disparity planes for the stereo pair with a baseline length of $B' = B+\Delta B_t$ = 50.51 mm. The zoomed iso-disparity planes with Z_n ranging from 1600 mm to 2400 mm are shown in Fig. 3b. The new iso-disparity plane is placed exactly in the middle, between the old iso-disparity planes Z_{49} and Z_{50} that correspond to a range of depths from 1997 mm to 1938 mm.

According to Eq. (11), for a target point T_0 with the disparity n_t = 48, the projection shift distance $D_a = 0.49\Delta$ and, as Fig. 2 shows, the dither signal amplitude $D_a = 0.5\Delta$, the optimum mean image horizontal quantization uncertainty interval is half a pixel, $\bar{q}_h = 0.5\Delta$. The change of the baseline length according to Eq. (12) places the new iso-disparity plane in the middle of the old iso-disparity planes and is thus optimal from a quantization point of view.

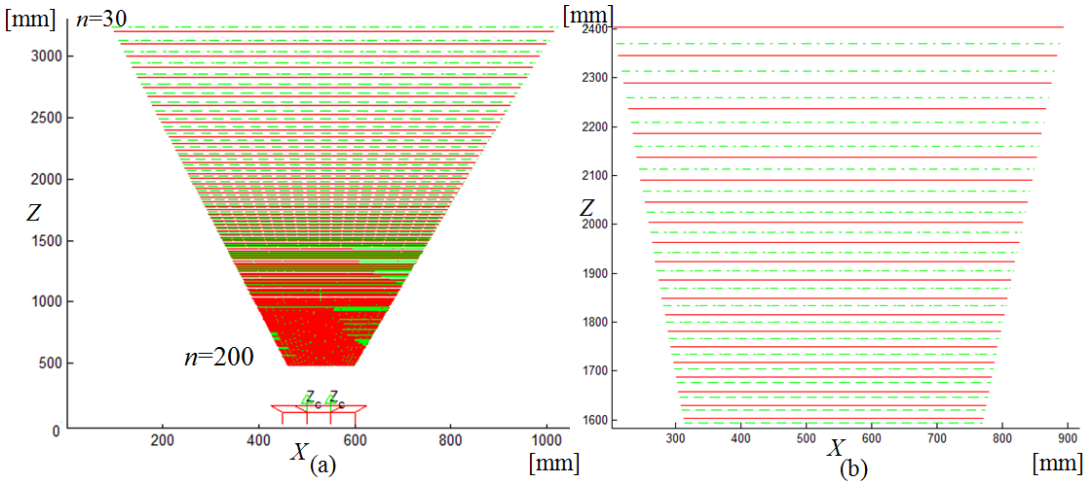


Fig. 3. (a) The iso-disparity planes for parallel cameras where Δ = 12.9 μm, and f = 25 mm; the red solid lines for B = 50mm, the green dashed lines for B= 50.51 mm. (b) The zoomed iso-disparity planes for Z_n were in the range from 1600 mm to 2400 mm.

3.5 The accepted 3D reconstruction space with defined accuracy

From Fig. 2, we know that the image horizontal quantization uncertainty can be reduced by half when the dither signal amplitude D_a is equal to the optimal value 0.5Δ, for a target point T_0 with the disparity, n_t. In order to generalize, we extend the target point to a bounded space defined by a cube. Here, we denote the deviation of the projection displacement as ΔD_a, and define it as the difference between the projections displacements of the target point T_0 and any point T on the stereo images with disparities, n_t and n respectively. It can be formulated as:

$$\Delta D_a(n) = |D_a(n) - D_a(n_t)| = \left| \frac{n - n_t}{2(n_t + 1)} \right| \cdot \Delta . \tag{13}$$

For the estimation of the depth range within a defined accuracy, we assume that the acceptable mean disparity quantization uncertainty $\bar{q}_d$ is $\pm 0.6\Delta$, which can be compared to the optimal $\pm 0.5\Delta$, then according to Eqs. (5) and (6) the mean image horizontal quantization uncertainty tolerance $\Delta\bar{q}_h$ is $\pm 0.3\Delta$, and from Fig. 2 the displacement D_a may be in the range of $[0.4\Delta, 0.6\Delta]$. In this case, the deviation of the projection displacement, ΔD_a, is 0.1Δ. The points in the accepted space with the disparity, $n_{\pm 0.3}(n_t)$, where ± 0.3 is the mean image horizontal quantization uncertainty tolerance $\Delta\bar{q}_h$, can be gotten from:

$$n_{\pm 0.3}(n_t) = \left\lfloor \pm \frac{2\Delta D_a}{\Delta}(n_t + 1) + n_t \right\rfloor = \lfloor \pm 0.2(n_t + 1) + n_t \rfloor, \tag{14}$$

where the floor $\lfloor \cdot \rfloor$ denotes the integer part. Furthermore, due to Eq. (11), the target point disparity n_t needs to be greater or equal to 4 in order for the D_a to be in the range of $[0.4\Delta, 0.6\Delta]$.

In Fig. 4, the red solid line shows the normalized deviation of the projection displacement, $|\Delta D_a/\Delta|$, versus the disparity, n, for three different target points with the disparities, $n_t = 20$, 49 and 176 respectively, the baseline length $B = 50$ mm, and the

focal length f = 25 mm. Table I shows the disparity ranges, $n_{\pm 0.3}(n_t)$, which correspond to the depth ranges $\Delta Z_{\pm 0.3}(n_t)$ for a target point with a disparity of n_t, which corresponds to the blue dashed lines in Fig. 4.

Table I. The disparity range and depth range of the points in the accepted space for the target point

n_t	Disparity range ($n_{\pm 0.3}(n_t)$)	Depth range ($\Delta Z_{\pm 0.3}(n_t)$)(mm)
20	15 - 24	6460 - 4037
49	39 - 59	2485 - 1642
176	140 - 211	692 - 459

4. Implementation

From Fig. 5, it can be seen that the dither signals d_{li} and d_{ri} control the position of the left and right cameras. The target point projections x_{li} and x_{ri} correspond to the i-th dither position of the left and right camera respectively and the quantized signals are x_{Qli} and x_{Qri} for the left and right image respectively. Furthermore, we can now calculate the target depth information by averaging the depths of all possible disparities d_i of the stereo pairs. The arithmetic average of all the depths constitutes an unbiased estimate of the target point depths and the depth reconstruction uncertainty is reduced by half for a two-stage discrete dither signal.

The dithering algorithm, when applying the two-stage discrete dither signal to the left and right cameras, can be divided into the four following steps:

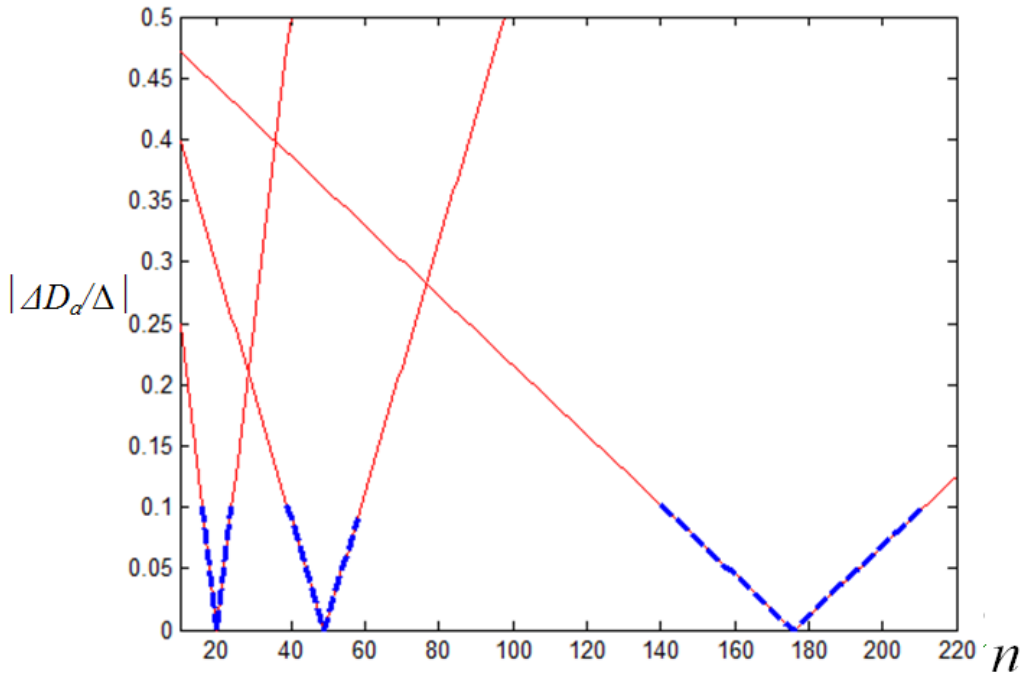


Fig. 4. The normalized deviation of the projection displacement, $|\Delta D_d/\Delta|$ represented by red solid lines, and $|\Delta D_d/\Delta| \leq 0.1$ represented by blue dashed lines; n_t=20, 49 and 179, while Δ= 12.9 μm, B=50 mm and f=25 mm.

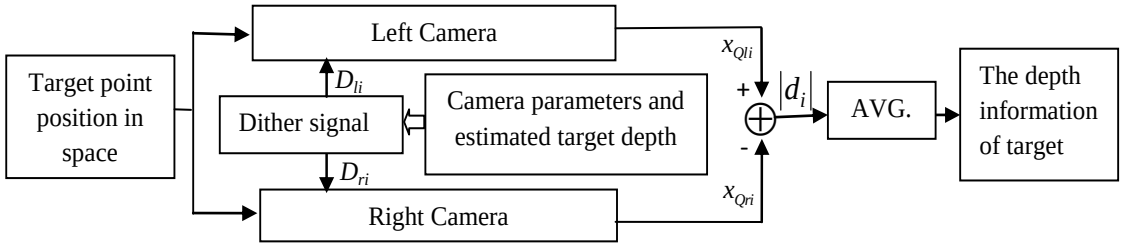


Fig. 5. The block diagram of the dithering algorithm where D_{li} and D_{ri} are the dither signals for the left and right cameras, x_{Qli} and x_{Qri} are the quantized signals for the left and right cameras, and d_i is disparity.

Step 1. The primary measurement of the target point depth where the target point is defined as a center of the target object:

The target point disparity and its depth can be estimated from Eq. (1) considering the stereo pair cameras' initial positions.

Step 2. Estimation and generation of the dither signal by the baseline length change:

According to Eq. (12), the baseline length change (increasing or decreasing) can be obtained from the estimated target point disparity. This change is also equal to the dither signal and can be used to control the left and right camera movement. For simplification of movement, the left and right cameras move in the same way and in the same direction.

Step 3. The secondary measurement and calculation of the new disparities:

We acquire the new left and right camera positions for the secondary measurement. It should be noted that now all four images have been taken; a first pair consisting of left and right images was taken before the cameras' movement and another pair of images was taken after the cameras' movement. Theoretically, we can combine those four images into six pairs of stereo images. In practice, however, the dither signal moves the cameras a very short distance and the disparity of the images taken by the camera pair at the two closest positions is too small for depth information to be usefully extracted from this pair. Thus, we have four pairs of images left that we can use to form the disparity matrix, **d**, from the horizontal image coordinate vectors $\boldsymbol{x}_{Ql}$ and $\boldsymbol{x}_{Qr}$:

$$\mathbf{d} = \left| \boldsymbol{x}_{Ql}{}^{\mathrm{T}} \cdot \boldsymbol{x}_{Qr} \right|, \tag{15}$$

where the matrixes $\boldsymbol{x}_{Ql}$ and $\boldsymbol{x}_{Qr}$ are extensions of the vectors $[x_{Ql1}\ x_{Ql2}]$ and $[x_{Qr1}\ x_{Qr2}]$ with the padding of -1 and 1 respectively, described as:

$$\boldsymbol{x}_{Ql} = \begin{bmatrix} x_{Ql1} & x_{Ql2} \\ -1 & -1 \end{bmatrix} \text{ and } \boldsymbol{x}_{Qr} = \begin{bmatrix} 1 & 1 \\ x_{Qr1} & x_{Qr2} \end{bmatrix},$$

where indexes 1 and 2 correspond to the positions of the initial and the new projections for the stereo pair cameras respectively, while T denotes a transpose of the matrix.

Step 4. The calculation of the final target point depth and its depth reconstruction quantization uncertainty:

The final target point depth can now be determined by averaging the depths from four pairs of images, which can be described as:

$$\bar{Z}_t = \frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{fB_{i,j}}{d_{i,j}}, \tag{16}$$

with

$$\mathbf{B} = \begin{bmatrix} B & B+\Delta B_t \\ B-\Delta B_t & B \end{bmatrix},$$

where B is the baseline for the stereo pair camera with its initial position, while ΔB_t is the length change of the baseline. The i and j are the row and column indexes in the matrix respectively.

5. Results

In this section, we describe two case studies where we test the mathematical models outlined above. The first study is a simulation experiment and the second is a physical experiment. The case studies both illustrate how the depth reconstruction uncertainty in stereo coverage is reduced by the dithering algorithm. The results of the studies are compared with a conventional direct triangular method, [23]. The simulation experiment was performed in MATLAB 7.0. The Epipolar Geometry Toolbox, [24], was used to project and transform the object position in 3D space. In both cases studies, the cameras' optical axes were parallel in the *XZ*-plane, the initial stereo pair baseline length was 50 mm, the focal length was 25 mm and the pixel length Δ = 12.9 μm.

5.1 Simulated statistical analysis of the depth reconstruction uncertainty

The simulation considered 1500 target points randomly located in a 300 mm×300 mm×300 mm cube. The cubic depth range was from 1800 mm to 2100 mm. The perspective view is shown in Fig. 6a. In the simulation, we applied a statistical analysis to the reconstruction uncertainty for the target points. We assumed that each point can be detected at least, and only, by one pixel on the image plane. The reconstructions of points were simulated with a direct triangular method and the result is shown as green points in Fig. 6b. The green points form iso-disparity planes which intervals illustrate the reconstruction uncertainty of the method. The red points correspond to the exact simulation points. To estimate the dither signal from the first measurement, the centers of the cubic disparity n_t = 49 was assumed to be the target point. From Eq. (12), we can determine that the left and right camera movement distance ΔB = 0.51 mm. All points in the cubic space are in a disparity range of 46 to 53, and according to Eq. (13) their maximum deviation projection displacement ΔD_a is 0.04Δ, which is inside the accepted space with a mean disparity quantization uncertainty of $\pm 0.6\Delta$. In Fig. 6c, the black points form the new iso-disparity planes reconstructed by the triangular method and making use of the dithering algorithm. The intervals of iso-disparity planes correspond to the depth reconstruction quantization uncertainty of the dithering method.

The histograms describing the normalized depth reconstruction errors for all the simulation points using both the direct triangular method and the dithering algorithm are shown in Fig. 7. The histograms show that the triangular distributions of the reconstruction uncertainty and their ranges are $[-\Delta, \Delta]$ and $[-0.5\Delta, 0.5\Delta]$ for the direct triangular method and the dithering algorithm respectively. The standard error deviation, σ, is 8.1 mm for the dithering algorithm and 16.1 mm for the direct triangular method. The improved depth reconstruction accuracy is 49.7 % in this case. This figure agrees with the theoretical analysis which proved that the depth reconstruction uncertainty should be reduced by half by the proposed algorithm.

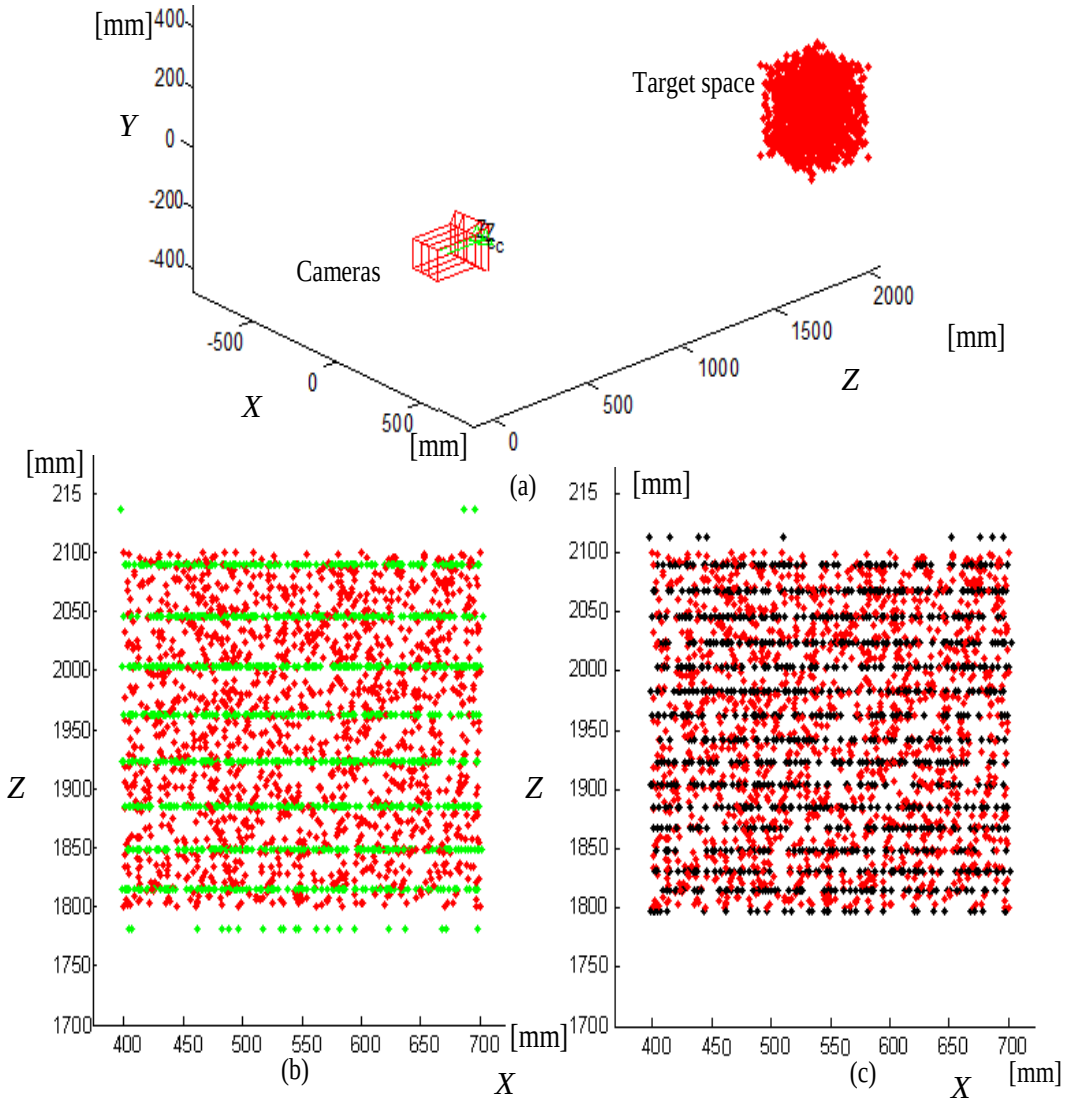


Fig. 6. Illustration of the first case study showing the exact points in red; (a) the perspective view of the cameras and the target points, (b) the direct triangular method showing the reconstructed points in green, and (c) the dithering algorithm showing the reconstructed points in black.

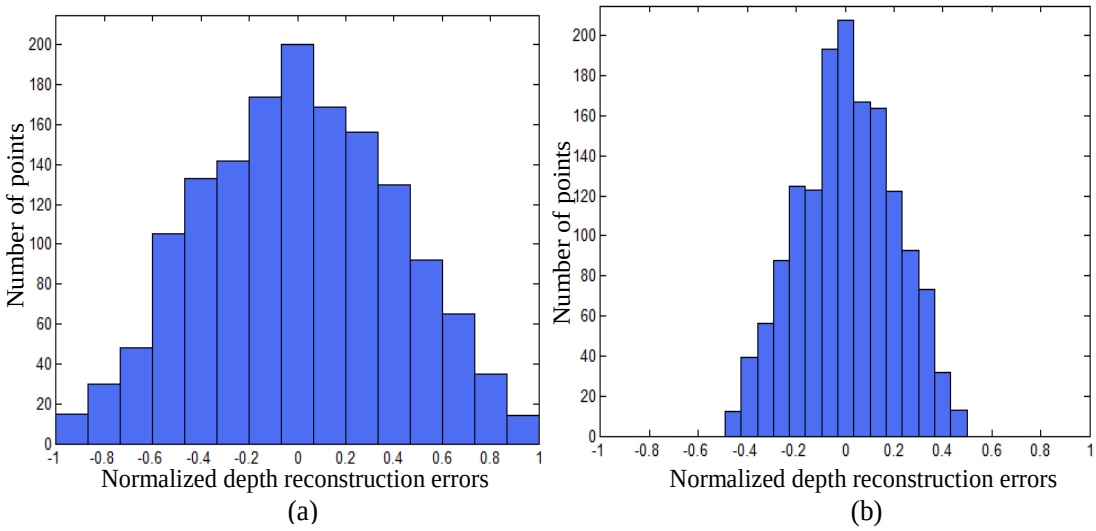


Fig.7. Histograms of the normalized depth reconstruction errors (a) for the direct triangular method and (b) the dithering algorithm.

5.2 Physical experiment

In practice, it is inconvenient to measure depth as the distance between the target object and the camera. It is preferable to measure a line differential depth, ΔZ_L. A line differential depth is defined as a distance between two target points along optical axis. The physical experiment was designed to measure a line differential depth of ΔZ_L using one camera and a linear translation stage. Fig. 8 shows a photo and the schematic diagram of the experimental stage, where the camera optical axis is oriented on and normal to the linear translation stage table. The origin of the coordinate system is located at the center of the lens and the linear translation stage table is parallel with the

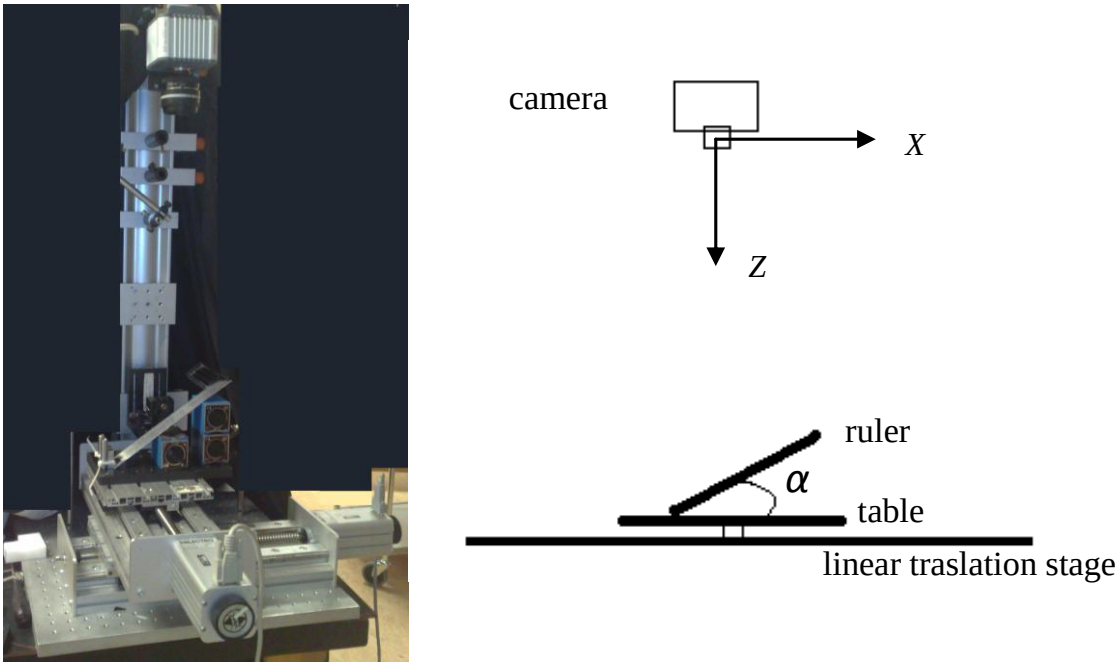


Fig.8. The photo and schematic diagram of objects and configurations in the physical experiment.

X-axis. A ruler functions as the target object and it is transported by the linear translation stage table along the *X*-axis with a resolution of 1/80 mm. The ruler is located about 550 mm away from the camera lens center and has the angle, α, to the table surface.

The camera focuses on the target points in the camera's depth of field. Two parallel lines on the ruler with the angle, α, are tested. The lines' lengths are 80 mm. Fig. 9 shows images of the example with the angle $\alpha_1 = 35.6^{\circ}$. The left and right images are captured for the ruler positions with a distance of 50 mm, which corresponds to a stereo pair with the base line length of 50 mm and the parallel cameras' optical axes. The minimal and maximal disparities of the points are 167 and 185 and their mean disparity value is 176. The maximum deviation projection displacement is 0.03Δ, which is inside the accepted space of the mean disparity quantization uncertainty $\pm 0.6\Delta$. To estimate the dither signal, a virtual point with the disparity 176 is assumed to be the target point. According to Eq. (12), the baseline should increase (decrease) by 0.14 mm. After shifting the ruler, we took the other left and right images. The ΔZ_L can be estimated by the dithering algorithm. The other two rulers' angles, $\alpha_2 = 19.7^{\circ}$ and $\alpha_3 = 0^{\circ}$, were also tested. Table II presents the results from a comparison between the direct triangulation method and the dithering algorithm. The exact line differential depth is taken from physical measurement and the reconstructed ones are from the stereo pairs' measurement using two different methods. Table II shows the calculated mean absolute errors to be 4.7 mm and 3.0 mm for the direct triangulation method and the dithering algorithm respectively. The improved depth reconstruction accuracy is thus 36.2 %.

Table II. Results of the reconstructed ΔZ_L showing errors by the triangulation method and dithering algorithm for the real target lines.

Angles α:*s* [°]	Line	Exact ΔZ_L [mm]	The direct triangulation method		The dithering algorithm	
			Reconstructed ΔZ_L [mm]	Reconstruction error [mm]	Reconstructed ΔZ_L [mm]	Reconstruction error [mm]
35.6	1	46.5	43.4	3.1	45.0	1.5
	2	46.5	41.9	4.6	43.4	3.1
19.7	1	27.0	21.9	5.1	23.6	3.4
	2	27.0	21.3	5.7	23.2	3.8
0	1	0	3.2	3.2	1.6	1.6
	2	0	6.4	6.4	4.8	4.8

6. Conclusion

One way of resolving the limitations of sensor sensitivity and image resolution due to SNR is to acquire a super-resolution image from a low resolution sensor. Through this process, it is also possible to reduce the spatial quantization caused by a discrete sensor. Super-resolution can be obtain by combining the information from a set of slightly different low-resolution images of the same scene. The dithering method can be applied to the model and control the movement of the cameras used to get these different images.

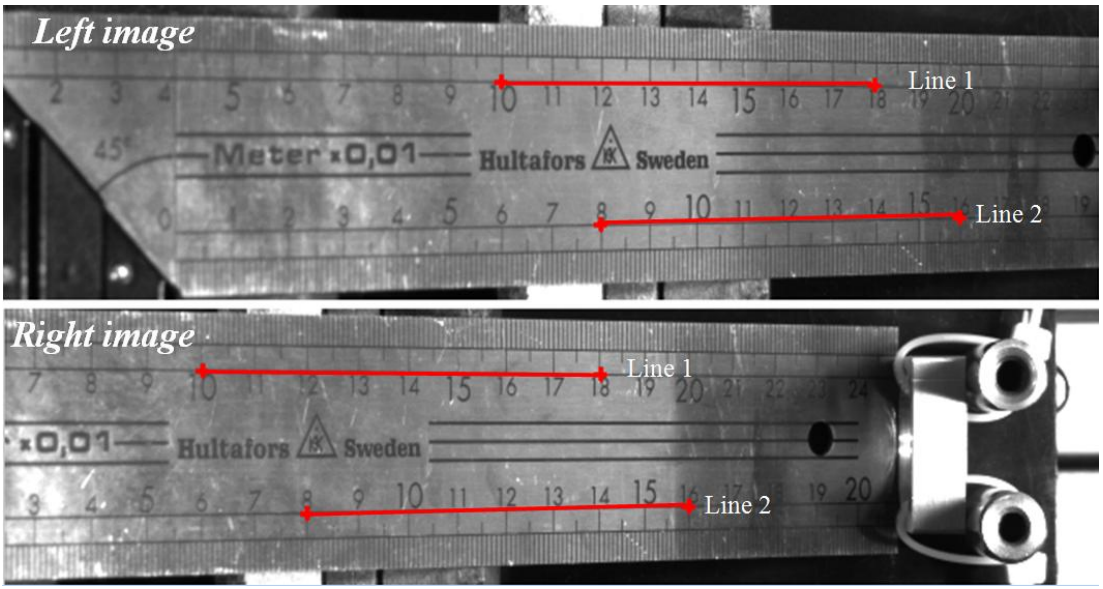


Fig.9. The images from the stereo pair of ruler with two parallel lines.

Analysis of the depth reconstruction uncertainty can be accomplished through the use of the mathematical model of depth reconstruction with dithering. There are two methods by which a dither signal can be added: the position of the target can be shifted or the position of the camera sensor can be shifted. The proposed method is based on movement of the camera sensor position. By applying the two-stage discrete dithering signal and combining four images into four pairs of stereo images, the depth reconstruction uncertainty can be reduced by half.

The discrete dither signals are estimated by analyzing the iso-disparity planes and then generated by controlling the stereo pair baseline length and placing the new iso-disparity plane exactly in the middle of the previous iso-disparity planes. This gives the optimal solution for controlling the camera movement.

The simple and robust dithering algorithm that significantly reduces the depth reconstruction quantization uncertainty is easy to implement through the use of the presented mathematical model. The proposed algorithm can also be used to position the stereo cameras. The dithering algorithm was verified through simulation and with the aid of a physical experiment. The results produced by the direct triangular algorithm and the dithering algorithm are contrasted and presented in the form of a statistical analysis. The histograms of the normalized depth reconstruction errors clearly show the triangular distribution of the reconstruction uncertainty, revealing that the target depth reconstruction uncertainty is reduced by half by the proposed algorithm. The result from the physical experiment proves that the line differential depth accuracy can be significantly improved by the dithering algorithm.

It should be taken into consideration that the real image can be distorted by other sources of noise, such as lens resolution or lens distortion. This is why the experimental results are slightly different from those of the theoretical analysis. Thus, the dither signal resolution is limited by the movement resolution which is a linear translation of the accuracy of the stage table. This affects the performance of the real system and limits the increase in accuracy.

It is not always necessary to build the entire depth information from the stereo images. If only a small target region is required, then the dithering algorithm could spend less computing time to build that region. For a low texture target region where the texture features are unclear, we can project a structured light pattern on the target to aid the correspondence algorithm to find accurate matches. For example, the pattern can be designed as dots that are projected onto the image plane within one pixel. Since the structured light pattern is known, it can be removed from the images to ensure a realistic, textured result. The applications could be used when the area demands very accurate depth reconstruction, for instance in human face feature extraction, remote controlled surgery etc. Furthermore, the measurement of the differential depth is also required in robotics controlling.

Further research is needed to explore this issue. In particular, it would be useful for 3D scene reconstruction to analyze the two dimensional image quantization uncertainty and its corresponding dithering methods. The dithering algorithm based on vergence angle changes by shifts of the camera sensor chip and the real-time implementation could be extended. Results from such research could not only be used for 3D reconstruction but would also aid the study of stereo matching algorithms. Future research could also focus on the reduction of the depth reconstruction uncertainty of the out-of-focus part of an image (the blurred part) or in cases where the target is blurred as a result of dynamic movement. Without the primary measurement of the target point depth, the dithering algorithm can still be implemented by shifting the sensor plane.

Acknowledgements

The authors wish to acknowledge Prof. Stefan Andersson-Engels, Pontus Svenmarker and Haiyan Xie at the Division of Atomic Physics, Lund University, Sweden for lending their expertise regarding the necessary laboratory equipment. The authors would also like to thank Dr. Benny Lövström at Blekinge Institute of Technology, Sweden and Dr. Fredrik Bergholm for valuable discussion and support. Finally, we would like to thank Dr. Johan Höglund for his comments.

References

[1] W. Kulesza, J. Chen, S. Khatibi, Arrangement of a Multi Stereo Visual Sensor System for a Human Activities Space, in: A. Bhatti (Ed.), Stereo Vision, InTech Education and Publishing, Vienna, 2008, pp. 153-172.

[2] J. Chen, S. Khatibi, and W. Kulesza, Planning of a Multi Stereo Visual Sensor System for a Human Activities Space, in: Proc. of 2nd International Conference on Computer Vision Theory and Applications, 2007, pp. 480-485.

[3] J. Chen, S. Khatibi, and W. Kulesza, Planning of a Multi Stereo Visual Sensor System Depth Accuracy and Variable Baseline Approach, in: Proc. of 3DTV Con. on the True Vision Capture, Transmission and Display of 3D Video, 2007.

[4] T. Chen, P. Catrysse, A. Gamal and B. Wandell, How Small Should Pixel Size Be? In: Proc. of SPIE on Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications, Vol. 3965, 2000.

[5] S. C. Park, M. K. Park and M.G. Kang, Super-Resolution Image Reconstruction: a Technical Overview, IEEE Signal Processing Magazine, 20(3) (2003), pp. 21-36.

[6] P. Vandewalle, S. Süsstrunk, and M. Vetterli, A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution, EURASIP Journal on Applied Signal Processing, 2006, pp. 1-14.

[7] A. Francisco and F. Bergholm, On the Importance of Being Asymmetric in Stereopsis-or Why We Should Use Skewed Parallel Cameras, International Journal of Computer Vision, 29(3) (1998), pp. 181-202.

[8] M. Ben-Ezra, A. Zomet, and S. K. Nayar, Video Super-Resolution Using Controlled Subpixel Detector Shifts, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(6) (2005), pp. 977-987.

[9] J. Chen, S. Khatibi, J. Wirandi and W. Kulesza, Planning of a Multi Stereo Visual Sensor System for a Human Activities Space – Aspects of Iso-disparity Surface, in: Proc. of SPIE in Electro-Optical Remote Sensing, Detection, and Photonic Technologies and Their Applications, Vol. 6739, 2007.

[10] R. A. Wannamaker, The Theory of Dithered Quantization, PhD thesis, the University of Waterloo, Canada, 2003.

[11] R. A. Wannamaker, S. P. Lipshitz, J. Vanderkooy and J. N. Wright, A Theory of Nonsubtractive Dither, IEEE Transactions on Signal Processing, 48(2) (2000), pp. 499-516.

[12] M. F. Wagdy, Effect of Various Dither Forms on Quantization Errors of Ideal A/D Converters, IEEE Transactions on Instrumentation and Measurement, 38(4) (1989), pp. 850-855.

[13] P. Carbone and D. Petri, Mean Value and Variance of Noisy Quantized Data, Elsevier Measurement, 23(1998), pp. 131-144.

[14] R. N. Hook, A. S. Fruchter, Dithering, Sampling and Image Reconstruction, in: Proc. of Astronomical Data Analysis Software and Systems IX Conference, Vol. 216, 2000.

[15] C. C. Yang and F.W. Ciarallo, Optimized Sensor Placement for Active Visual

Inspection, Journal of Robotic Systems 18(1) (2001), pp. 1-15.

[16] X. Liu and R. W. Ehrich, Subpixle Edge Location in Binary Images Using Dithering, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(6) (1995), pp. 629-634.

[17] W. N. Klarquist and A. C. Bovik, FOVEA: A Foveated Vergent Active Stereo Vision System for Dynamic Three-Dimensional Scene Recovery, IEEE Trans. on Robotics and Automation, 14(5) (1998), pp. 755-770.

[18] A. Basu and H. Sahabi, Optimal Non-uniform Discretization for Stereo Reconstruction, in: Proc. of the 13th International Conference on Pattern Recognition, 1996, pp. 755-759.

[19] Y. J. Kil, B. Mederos and N. Amenta, Laser Scanner Super-resolution, in: Eurographics Symposium on Point-Based Graphics, 2006.

[20] H. Farid, Range Estimation by Optical Differentiation, PhD Dissertation, the University of Pennsylvania, 1997.

[21] P. Carbone and D. Petri, Effect of Additive Dither on the Resolution of Ideal Quantizers, IEEE Transactions on Instrumentation and Measurement, 43(3) (1994), pp. 389 - 396.

[22] J. J. Rodríguez and J. K. Aggarwal, Quantization Error in Stereo Imaging, in: Proc. of Computer Vision and Pattern Recognition, 1988, pp. 153-158.

[23] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Second Edition, Cambridge University Press, 2004.

[24] G. L. Mariottini and D. Prattichizzo, EGT for Multiple View Geometry and Visual Servoing: Robotics Vision with Pinhole and Panoramic Cameras, IEEE Robotics & Automation Magazine, 12(4) (2005), pp. 26-39.

Paper III

Appling Dithering to Improve Depth Measurement Using a Sensor-Shifted Stereo Camera

Authors:

Jiandan Chen, Wail Mustafa, Abu Bakr Siddig and Wlodek Kulesza

Reformatted version of paper originally published in:

Metrology and Measurement Systems, vol. 17, no. 3, pp. 335-348, October, 2010.

Appling Dithering to Improve Depth Measurement Using a Sensor-Shifted Stereo Camera

Jiandan Chen, Wail Mustafa, Abu Bakr Siddig and Wlodek Kulesza

Abstract

The sensor-shifted stereo camera provides the mechanism for obtaining 3D information in a wide field of view. This novel kind of stereo requires a simpler matching process in comparison to convergence stereo. In addition to this, the uncertainty of depth estimation of a target point in 3D space is defined by the spatial quantization caused by the digital images. The dithering approach is a way to reduce the depth reconstruction uncertainty through a controlled adjustment of the stereo parameters that shift the spatial quantization levels. In this paper, a mathematical model that relates the stereo setup parameters to the iso-disparities is developed and used for depth estimation. The enhancement of the depth measurement accuracy for this kind of stereo through applying the dithering method is verified by simulation and physical experiment. For the verification, the uncertainty of the depth measurement using dithering is compared with the uncertainty produced by the direct triangulation method. A 49% improvement of the uncertainly in the depth reconstruction is proved.

Keywords**:** *Depth Reconstruction, Dithering, Skewed Parallel, Stereo Setup, Iso-disparity Surfaces.*

1. Introduction

Advanced technologies may help people to extend their ability to process visual information. This raises the demand for autonomous systems with high performance sensors. This paper is concerned with a stereo camera and its applications in human activity monitoring. The Intelligent Vision Agent System, IVAS, is a vision and information processing system used in these kinds of applications, [1, 2] and [3]. The IVAS gathers data in order to reconstruct 3D information that can be used in health care, security and surveillance applications. The system focuses on an interesting part of the scene by dynamic control of the stereo image. Such a system requires high accuracy in the reconstruction of the 3D information in order to guarantee high performance.

The human activity field is a 3D world, where the location of each point is represented by *x*, *y* and *z* coordinates. However, a camera can only capture a two-dimensional image where each point is represented by *x* and *y* coordinates. The stereo system provides the mechanism for acquiring the vital *z* coordinate, [4]. In this case, the *z* coordinate is referred to as the depth, and the process of acquiring the depth from the stereo system is called depth reconstruction.

The stereo system captures two images in the 3D world. The reconstructed point needs to have a projection in each image, and this can only happen when the point is located in the common field of view (FoV) of the camera pair. To be able to reconstruct 3D information, the system needs to first implement a matching process to find the corresponding points in the two images of the same view.

However, a digital camera quantizes the image plane into an array of pixels that forms the digital image. Because of this, the projection points are approximated and assumed to be located in the centers of the pixels. The difference between the reconstructed depth estimated from the exact and the discretised projections is referred to as the depth reconstruction uncertainty. The depth reconstruction uncertainty is related to the pixel size of the camera sensor. The selection of an optimal sensor pixel size is discussed in [5].

Regarding the image resolution, the limitation of the pixel size is overcome by combining the information from slightly different low-resolution images of the same scene into a higher-resolution image. This way of enhancing the image resolution is called super-resolution reconstruction, [6]. In a similar way, two pairs of images taken by a stereo system with two slightly different setups can be used in combination to reconstruct the depth with enhanced uncertainty. The two different setups can result from a readjustment of one of the stereo parameters. This method is referred to as the dithering approach, [7]. The dithering technique is commonly used to overcome the true color issue that exists in colored image printing, [8]. The effect of the dithering on the estimation of the sine wave amplitude is studied in [9].

In many cases, the camera has to be rotated to capture a certain view that is initially outside the camera FoV. This rotation causes distortion of the shapes in the captured images. Hence, further processing is required to overcome this distortion. Instead of rotating the whole camera, professional photographers use a technique called a sensor-shifted camera that applies a millimeter shift between the camera lens and the sensor, [10]. This shift provides an effect similar to the rotation, so that it captures the wanted view while avoiding the distortion.

Francisco and Bergholm have proposed the use of a sensor-shifted camera in the stereo setup where the sensor has a controlled micro-movement, [11]. This kind of stereo system is referred to as the skewed-parallel stereo camera. In that paper, the authors discuss the benefits of using this kind of camera in a stereo system instead of the vergence movement used in the general stereo setup. When compared to the convergence camera stereo setup, the skewed-parallel camera setup requires simpler reconstruction processing. In addition to this, the skewed-parallel camera setup offers a wider common field of view than the parallel camera stereo setup. A similar camera was used by Ben-Ezra et al. to minimize the motion blur in the reconstruction of super-resolution images of a video signal, [12].

After an introduction and a summary of the related works, the problem statement and the main contributions are provided in Section 2. In Section 3, the geometric model of the skewed-parallel stereo camera is described followed by the derivation of the dither signal for this stereo system. Section 4 discusses the implementation of the dithering algorithm along with the conducted synthetic and physical validation experiments, and their results. The prototype of the skewed-parallel camera used in the physical

experiment is also described in this section. Finally, a conclusion and a recommendation for future work are provided in Section 5.

2. Problem statement and main contributions

The concern of this paper is to improve the depth measurement uncertainty through the use of the dithering algorithm for a specific setup of the skewed-parallel stereo camera. The uncertainty of the depth measurement can also be reduced by decreasing the pixel size but the drawback of this method is that this will also reduce the signal-to-noise ratio. Therefore, we propose the application of the dithering method as a simple and robust way to improve the reconstruction uncertainty. The modeling of the depth reconstruction using the dithering algorithm for this kind of stereo system, and its validation through simulation and physical experiment, are the main problems that need to be resolved before applying the method to real measurement.

This paper contributes to current research by:

- Developing a mathematical model of the depth measurement using the skewed parallel stereo system.
- Developing and implementing the dithering algorithm based on the depth measurement model.
- Configuring a prototype for the skewed-parallel camera setup.
- Validating the depth reconstruction enhancement of the dithering algorithm for the skewed-parallel camera stereo setup through simulation and physical experiment.

3. Problem analysis and modelling

3.1 The skewed-parallel stereo geometric model

In our approach, we use the pinhole camera model. The setup of the skewed-parallel stereo camera in the x–z plane is shown in Fig. 1, [13]. The center of the coordinates is in the middle of the baseline, B. The baseline is the distance between the optical lenses' centers, o_l and o_r, of the left and the right cameras respectively. The sensors lie on the same horizontal line, and the sensor centers are denoted as c_l and c_r for the left and the right cameras respectively. The shift of the sensor is defined as the horizontal distance between the optical center and the sensor center and it can be different for each camera. The shifts of the sensors of the left and the right cameras are denoted as S_l and S_r respectively. In our notation, the shifts are positive for a movement of the sensors to the right, and negative for an opposite direction. The focal lengths of the two cameras are assumed to be the same and are denoted as f.

The angle between the optical axis and the primary axis, which is defined as the line passing through the sensor center and the center of the lens, is called the convergence angle and it is denoted as α_l and α_r for the left and right cameras respectively. The convergence angle exists as a result of shifting the sensors of the skewed-parallel stereo cameras. This shift has the same effect as rotating the stereo pair since it introduces a fixation point, P_0, and widens the common FoV. Using trigonometry, the convergence angle of the left camera can be derived as:

$$\alpha_l = \tan^{-1}\frac{S_l}{f} \quad \text{for} \quad -m/2 \le S_l \le m/2 \ , \tag{1}$$

where m is the length of the sensor plane.

For any point in the space, P, with the depth Z, the projections of the point on the left and right sensors along the x-axis are x_l and x_r respectively (the x coordinates of each sensor cross the sensor middle). Considering the quantization effect, these projections are approximated by the pixel centers, and denoted as x_{Ql} and x_{Qr} for the left and right cameras respectively. In this case, the quantized depth Z_q of the point can be found through:

$$Z_q(n) = \frac{fB}{(n\Delta D + (S_r - S_l))} \quad \text{for} \quad n\Delta D + (S_r - S_l) > 0 \ , \tag{2}$$

with

$$n = x_{Qr} - x_{Ql} = \left\lfloor \frac{x_r}{\Delta D} \right\rceil - \left\lfloor \frac{x_l}{\Delta D} \right\rceil \ , \tag{3}$$

where ΔD is the length of pixel, n is an integer number representing the target disparity, and the symbol $\lfloor \cdot \rceil$ denotes rounding to the nearest integer.

The depth Z_0 of the fixation point, P_0, being a cross section of the primary axes, can be found by setting n to zero in Eq. (2). The x and z coordinates of the fixation point are:

$$\left(X_{P_0}, Z_{P_0}\right) = \left(\frac{S_r + S_l}{2(S_r - S_l)}B, \frac{fB}{S_r - S_l}\right) . \tag{4}$$

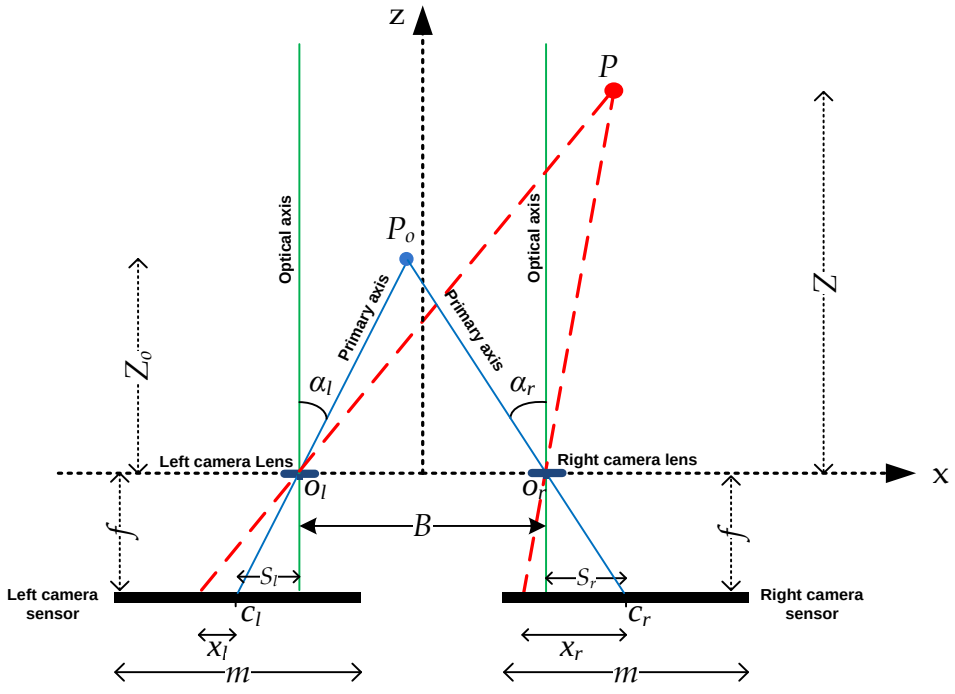


Fig. 1. An x-z view of the skewed-parallel stereo camera schematic diagram.

The disparity is the displacement of the corresponding *x*-axis projections of a certain point in space on the left and the right images. From equation (2), it can be seen that the depth reconstruction is inversely proportional to the disparity. For each disparity *n*, there exists a corresponding iso-disparity surface that represents the depth of all the points that have the same disparity, [7]. The iso-disparity surfaces appear as lines in the *x*–*z* plane as seen in Fig. 2 for the red lines. The interval between the iso-disparity surfaces represents the depth reconstruction quantization uncertainty, which is a nonlinear function of *n*.

3.2 Application of the dither signal

The idea behind the dithering technique is to add noise to the signal prior to the quantization process in order to slightly change the statistical properties of the quantization, [2]. The quantizers in this model are the cameras and the quantized signals are the target point projections x_l and x_r for the left and right cameras respectively. Using the dithering technique to reduce the uncertainty of the depth reconstruction for the parallel stereo setup was proposed in [7]. In that paper, authors proved that the uncertainty is reduced by half by applying the two stage discrete binary dither signal. To accomplish this reduction, a dither signal adjusts the stereo setup for a secondary measurement that follows the initial one, and the depth can then be estimated from all these measurements.

In this paper, we use a two-stage discrete binary dither signal for each camera. This means that we make use of four images to calculate the depth of the target. This allows us to estimate the depth with a reduced quantization uncertainty. In the parallel stereo setup, the depth reconstruction uncertainty is halved when the dithering algorithm is applied. The optimal diter signal makes the target projection move from its original position by a distance that is equal to half a pixel size, [7]. This means that the dither signal shifts the iso-disparity line so that it lies in the middle of the two consecutive iso-disparity lines n_t and n_t+1 between which the target is present. This shift reduces the uncertainty by half.

For the skewed-parallel stereo camera setup, calling (2), the difference between two consecutive iso-disparity lines, ΔZ_t, which represents the depth reconstruction uncertainty, can be found to be:

$$\Delta Z_t = \frac{Bf\Delta D}{\left[n_t\Delta D + \left(S_r - S_l\right)\right]\left[\left(n_t + 1\right)\Delta D + \left(S_r - S_l\right)\right]} \quad , \tag{5}$$

where *t* refers to a specific iso-disparity line n_t.

To estimate the dither signal, generated by a shift of one sensor, that affects the iso-disparity line n_t to move exactly into the middle of ΔZ_t, we should first determine the depth in the middle by:

$$Z_t + \frac{\Delta Z_t}{2} = \frac{Bf}{n_t\Delta D + \left(S_r - S_l\right) + \Delta S_t} \quad , \tag{6}$$

where ΔS_t is the shift introduced to one sensor that moves the iso-disparity surfaces to the middle of the two consecutive iso-disparity lines n_t and n_t+1.

Calling (2) and (5), the dither signal, ΔS_t, can be mathematically proven to be:

$$\Delta S_t = -\frac{\left(n_t \Delta D + \left(S_r - S_l\right)\right)\Delta D}{2\left(\left(n_t + 1\right)\Delta D + \left(S_r - S_l\right)\right) + \Delta D} \quad . \tag{7}$$

To verify the dither signal obtained by (7), Matlab 7, [14], and the Epipolar Geometry Toolbox, [15], are used. In the verification scenario, a target in the common FoV that belongs to a specific iso-disparity line n_t is chosen. Then, the dither signal is applied to check the movement of the iso-disparity lines. For the simulation, the baseline B is set to 100 mm, the focal lengths f are 25 mm each, the pixel length ΔD is 8.33 µm and the target is assumed to be in the disparity line n_t = 178.

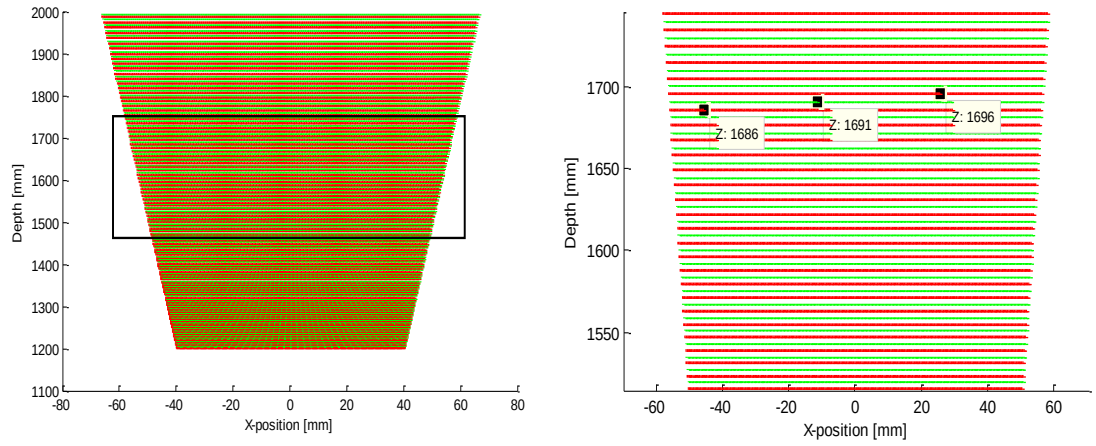
Fig. 2a shows the iso-disparity lines in 2D. Red lines denote the iso-disparity lines between 1200 mm and 2000 mm of depth for the original setup, while the green lines denote the iso-disparities after introducing the dither signal, ΔS_t, to both cameras. In this case, ΔS_t is found to be -4 µm approximately, which is equal to 0.5 pixels. This shows that the skewed-parallel stereo camera has the same property as the parallel stereo camera regarding the dither signal, [7]. To place the new iso-disparity surfaces in the middle of the old ones, it is required that the projection of the target feature is shifted half a pixel.

Fig. 2b shows a zoomed area around n_t = 178. From calculation, before applying the dither signal, the depth for the specified disparity can be determined to 1686 mm and 1696 mm for the next disparity level. After applying the dither signal, the depth for the same disparity n_t = 178, is found to be 1691 mm, which falls in the middle between the above-mentioned depths.

4. Implementation and validation

4.1 Implementation of the dithering algorithm

From the description of the dithering algorithm in [7], and the description of the estimation of the dither signal for the skewed-parallel stereo camera in section 3.2, the dithering algorithm can be defined. By applying the dither signal ΔS_t which controls the sensor positions of the left and right cameras, four images are obtained; two images



a) b)

Fig. 2. Iso-disparities obtained by simulation before and after applying the dither signal. Red lines come from the primary and green lines were obtained after the dither signal. b) Zoomed-in area (inside the box) of a).

before dithering and two images after dithering. For the disparity calculations, we can combine those four images into six pairs of images. In practice, however, since the dither signal moves the camera sensor a very short distance and the disparity of the two pairs of images taken by the same camera is too small to be useful to extract depth information from this pair. Therefore, only four pairs are considered.

The quantized projections of the target point obtained from the right and left images can be used to create the projection matrices $\boldsymbol{x}_{Qr}$ and $\boldsymbol{x}_{Ql}$ for the right and left cameras respectively as follows:

$$\boldsymbol{x}_{Qr} = \begin{bmatrix} x_{Qr1} & x_{Qr2} \\ -1 & -1 \end{bmatrix} \text{ and } \boldsymbol{x}_{Ql} = \begin{bmatrix} 1 & 1 \\ x_{Ql1} & x_{Ql2} \end{bmatrix} , \tag{8}$$

where x_{Qr1} and x_{Qr2} are the quantized projections of the target on the right image before and after applying the dither signal respectively, while x_{Ql1} and x_{Ql2} are the respective projections on the left image.

From the projection matrices, the dithering matrix, that contains the disparities of the four considered pairs, can be obtained by:

$$\mathbf{d} = \boldsymbol{x}_{Qr}{}^{\mathrm{T}} . \boldsymbol{x}_{Ql} \quad . \tag{9}$$

The dithering algorithm can be then implemented through the four following steps:

1. Preliminary estimation of the depth from the disparity of the initial pair of images using equation (2).
2. Estimation and application of the dither signal, ΔS_t, (7) to shift the sensors of the two cameras.
3. Secondary estimation of the four depths corresponding to the four disparities in **d**, equation (9).
4. Calculation of the depth of the target point by averaging the four depths from step 3.

4.2 Synthetic experiment

The synthetic experiment was performed using Matlab 7 and the Epipolar Geometry Toolbox, [15]. The simulation environment is a 3D space, with two pinhole skewed parallel cameras. The target is assumed to be 1500 points randomly distributed in a cubic area with the dimension 300 mm × 300 mm × 300 mm. The cubic center is set to be (0, 0, 1600) in *XYZ* coordinates. The target points fall in the common FoV of the stereo camera. Fig. 3 shows a perspective view for this setup in 3D space.

The setup of the stereo camera is: the baseline B is 100 mm, the focal length f is 25 mm, and the pixel size ΔD is 8.33 μm. The simulation scenario is to measure the depth of each point, where the initial shifts of the two cameras are set to zero, using the direct method through equation (2), and using the dithering algorithm described in Section 4.1. The results present a comparison between the two methods in order to illustrate how the depth reconstruction uncertainty can be improved by the dithering algorithm for this kind of stereo system.

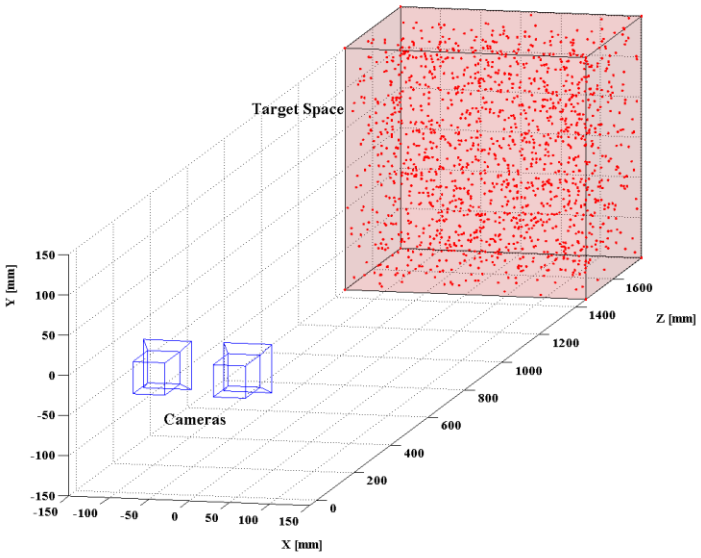
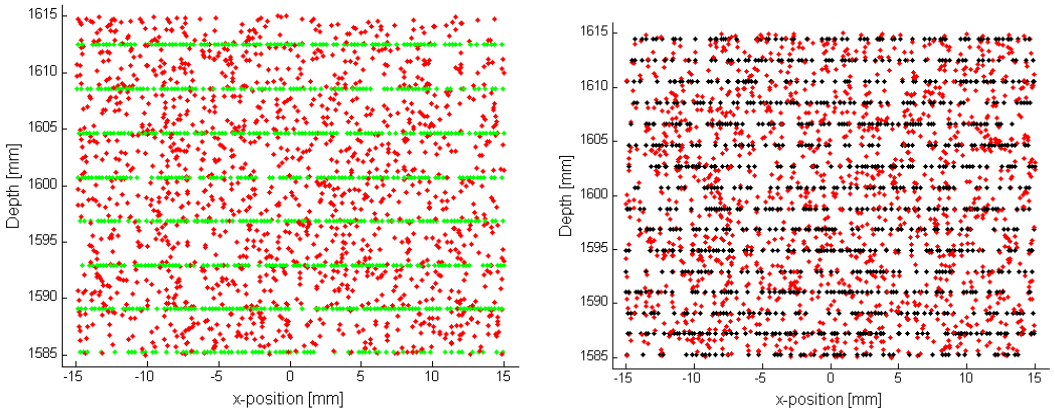


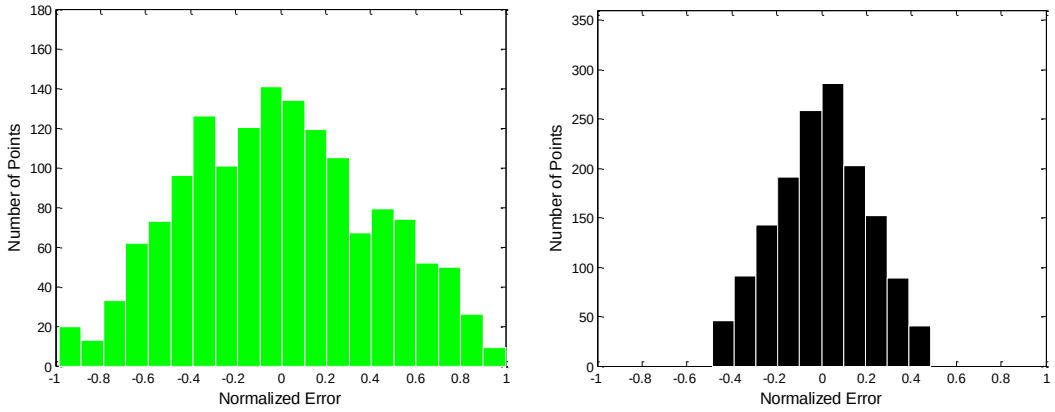
Fig. 3. The simulation setup: The points in the cubic area (red dots) are the targets.

Fig. 4 shows the top view of the target points in their original positions and in their reconstructed positions after the depth measurement for the two methods. The figure shows the points within the zoomed range from -15 mm to 15 mm along the *X*-axis, and from 1585 mm to 1615 mm along the *Z*-axis. In Fig. 4a, the red dots represent the original target points while the green dots represent the depth estimations for these points using the direct reconstruction method. The black dots in Fig. 4b represent the depth estimations for the same target points using the dithering algorithm. The black dots form new iso-disparity surfaces with different intervals that correspond to the reconstruction uncertainty of the method when applying the dithering algorithm.



a) b)

Fig. 4. Top view of the depth reconstruction of the original target points (in red). The green points represent the reconstructed points by the direct method. b) describes the reconstructed points by the dithering algorithm as black points.



a) b)

Fig. 5. Histograms of the normalized depth reconstruction error a) for the direct triangular method and b) after applying the dithering algorithm.

4.3 Skewed-parallel Camera Prototype

To physically implement the dithering algorithm for the skewed-parallel camera, it is necessary to make use of a camera with a sensor that is capable of moving in a controlled horizontal movement in relation to the lens. To satisfy this requirement, a prototype can be built by separating the camera body, which contains the sensor, from the lens that is normally attached to that body. Additionally, to control the movement of the camera sensor, the camera body needs to be attached to a micro-movement mechanical device such as an *x*-positioner or a DC motor.

Following the above design considerations, a skewed parallel camera prototype was built to be used for the experimental needs, [11]. This prototype was reconfigured for the purpose of this project. The camera prototype contains four main components that are: a camera module, a linear stage controlled by a DC motor, a lens and a metal stand that holds all the components. In Fig. 6, the prototype containing these components along with a target grid pattern that is used for the experiment part of the paper are shown.

The camera module is connected through a cable to a frame grabber card that is installed in a computer where the pictures are stored and processed. This module contains a CCD sensor that represents the shifted sensor in the skewed-parallel camera model. The Sony XC-555P camera module is used in this prototype. The camera module is a color video camera with a 1/2” type sensor, [16]. The captured image resolution is 768 (H) x 576 (V) pixels.

In order to provide the horizontal shift of the camera module, the prototype uses a linear stage with a DC motor. The linear stage movement distance is up to 25 mm with a resolution of 0.06 μm/count (motor step), [17]. This resolution is sufficient to provide the required micro-movement capability of the sensor.

To configure this camera prototype as a stereo camera, it has to be placed in two different positions where the distance between the two positions represents the baseline *B* of the stereo camera. The prototype has been attached to a translational stage with an accuracy of 1 mm, allowing an accurate implementation of the position.

The Tamron 23FM25SP lens model is used in this prototype. The lens is a C-mount type with a focal length of 25 mm. The focus of the lens can be adjusted for objects that lie between 0.15 m and ∞ from the front of the lens, [18].

4.4 Physical experiment

To validate that the dithering algorithm enhances the depth reconstruction for the skewed-parallel camera, we designed a physical experiment, [19]. The setup of the experiment is shown in Fig. 6 whereas the camera prototype is described in Section 4.3. The targets have been represented as points in a pattern consisting of a grid of lines with a distance of 5 mm between the lines. The pattern has been pasted onto a board that can be easily positioned in front of the stereo camera along the optical axis.

In [7], a way to validate the enhancement of the depth reconstruction is proposed. The idea is to measure the differential depth Z_{AB}, which is the distance between two targets along the optical axis, instead of the absolute depth. The method provides more accurate validation because it avoids measuring the distance between the test target and the center of the lens, something which is difficult to determine.

The reconstruction uncertainty of the differential depth, ΔZ_{AB}, represents a difference between the reference value of Z_{AB} and the reconstructed value by the stereo vision system. The probability distribution function, PDF, of the differential depth quantization uncertainty is described as the convolution of the depth quantization uncertainties at the two target points:

$$p(\Delta Z_{AB}) = p(\Delta Z_A) \otimes p(\Delta Z_B) \ , \tag{10}$$

where ΔZ_A and ΔZ_B are depth reconstruction quantization uncertainties at the target points A and B respectively and $\otimes$ denotes convolution. The PDF of the depth uncertainty ΔZ is defined as [20]:

$$p(\Delta Z) = \begin{cases} \dfrac{n_t^2 \Delta D}{(Bf)^2}\left(\Delta Z n_t^2 \Delta D + Bf\right), & -\Delta Z_m \le \Delta Z < 0 \\ -\dfrac{n_t^2 \Delta D}{(Bf)^2}\left(\Delta Z n_t^2 \Delta D - Bf\right), & 0 \le \Delta Z \le \Delta Z_m \\ 0, & \text{elsewhere} \end{cases} \ , \tag{11}$$

where ΔZ_m is the maximum depth reconstruction uncertainty corresponding the to the interval between iso-disparity surfaces.

According to (10), the range of the differential depth reconstruction quantization uncertainty is the sum of the depth quantization uncertainty ranges of the two corresponding points.

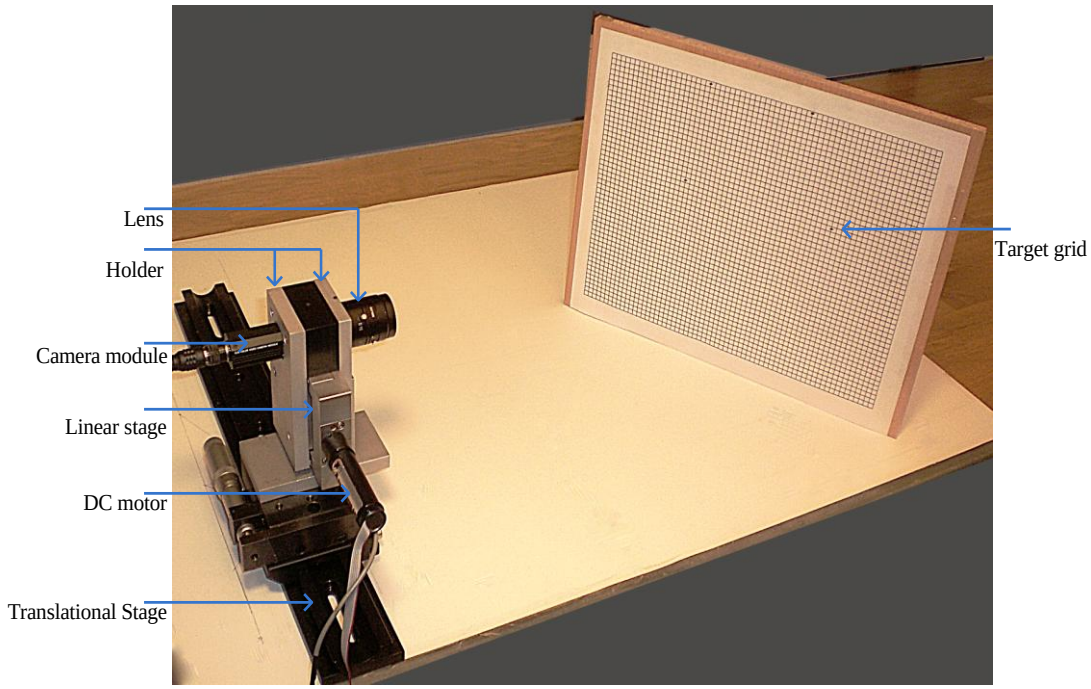
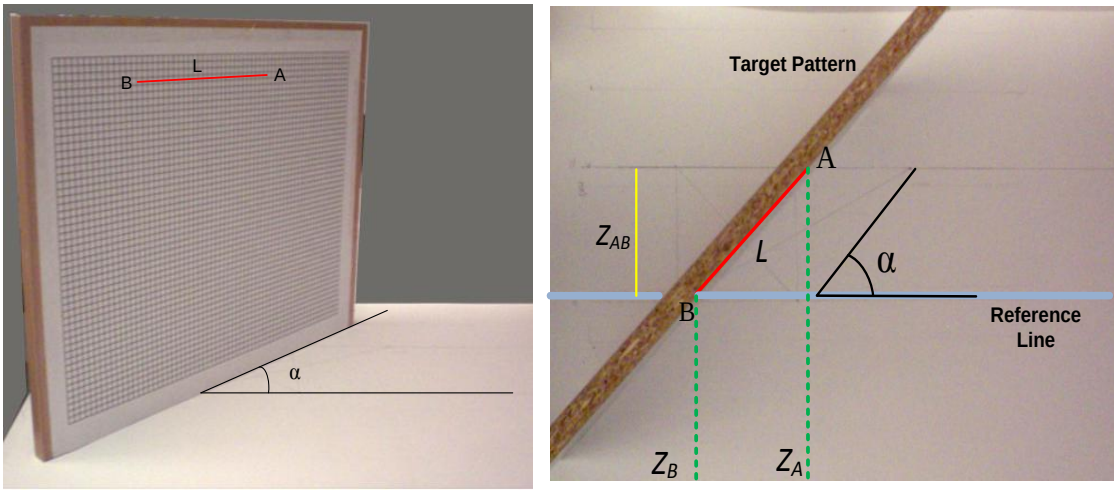


Fig. 6. A side view of the camera prototype showing its components and the experiment setup.

The experiment was conducted for three different values of the tilting angle α: 0°, 26.6° and 45°. In addition to this, two target pairs were used. The distance between the two target points on the grid *L* was set to 100 mm and 150 mm for the first and the second pair, respectively. The two target pairs were in the common field of view of the stereo system.

The baseline *B* is equal to 100 mm, and the grid of targets is at a distance of approximately 1600 mm from the baseline. For the camera prototype described in 4.3, the length of the pixel size has been calibrated to 8.33 µm, and the number of motor steps required to shift the camera module half a pixel size is found to be 69 motor steps. The focal length is 25 mm.



a) b)

Fig. 7. The target grid positioning: a) front view, b) top view.

Table 1 presents the results of the validation experiment for the two pairs of target points with the three different titling angles. The reference differential depth for each pair at each angle is compared with the reconstructed one by use of both the direct method and the dithering algorithm. The absolute error of the differential depth reconstruction for each method with respect to the reference differential depth is also listed for each case. This error can be estimated as a convolution of quantization errors of the target pair where these errors depend on the distance of each target point from the baseline. The distances cannot be measured exactly but it is possible to determine the minimum distance value to be 1560 mm. From this measure, it is possible to estimate that the quantization error of the target pairs is at least 4.7 mm which corresponds to $n = 253$. Then, the maximum absolute error is at least 9.4 mm for the reconstructed differential depth by the direct method in this setup, and 4.7 mm for the reconstructed differential depth by the dithering method.

From Table I, it can be noticed that the reconstruction error by the use of the dithering method is about half of that of the direct method. The mean of the absolute reconstruction errors for each method, calculated from the table, is 5.1 mm for the direct method, and 2.6 mm for the dithering algorithm. The improvement in the depth reconstruction accuracy in this experiment is thus 49%.

Table I. Results of the differential depth reconstruction showing errors by the direct method and the dithering algorithm for the target pairs.

Angles α:s [°]	Line length [mm]	Reference differential depth distance Z [mm]	The direct reconstruction method		The dithering method	
			Reconstructed Z_d [mm]	Absolute reconstruction error [mm]	Reconstructed Z_m [mm]	Absolute reconstruction error [mm]
0	100.0	0.0	7.4	7.4	3.7	3.7
	150.0	0.0	7.3	7.3	3.7	3.7
26.6	100.0	44.7	47.7	2.7	43.5	1.2
	150.0	67.1	61.0	6.1	63.9	3.2
45	100.0	70.7	68.8	2.0	71.7	1.0
	150.0	106.1	111.2	5.1	109.9	2.8

5. Conclusion

This paper introduces the use of sensor-shifted cameras in the stereo system instead of the conventional cameras and applies the dithering algorithm to improve depth reconstruction. Both the synthetic and the physical experiments verify that applying dithering reduces the depth reconstruction uncertainty by half when compared with the direct method.

It is verified by simulation that the dither signal causes the iso-disparity surfaces of the skewed-parallel stereo camera to shift with respect to their initial positions. The new position is the middle of the iso-disparity intervals before applying the dither signal. This can be interpreted to mean that dithering can be applied to reduce the depth

reconstruction uncertainty by half through a multi-stage measurement, i.e. the dithering algorithm.

It is also found that the dither signal is equal to half the pixel size of the camera sensor regardless of the first estimation of the depth of the target. Thus, it can be concluded that no primary measurement is needed for the dithering algorithm, which in turn means that it significantly simplifies the use of this kind of camera when determining depth.

In the synthetic experiment, the enhancement was verified using computer simulation. The depth reconstruction of the test target points using the dithering algorithm shows improvement of the uncertainty since the reconstruction of the points forms iso-disparity lines with interval widths reduced by half for the dithering algorithm when compared to the direct method. This improvement can be confirmed by comparing the histograms of the reconstruction error produced by both methods. This comparison shows that the dithering algorithm reduces the span of the error distribution to half the range obtained with the direct method. The depth reconstruction improvement can also be observed in the 48.6% reduction in the standard deviation of the reconstruction error for the simulation targets by the dithering algorithm.

The results of the physical experiment also show reconstruction improvement for all the test targets with an average of 49%, which is close to the theoretical value. For higher experimental accuracy, the differential depth of the target pairs was used in the validation experiment instead of the absolute depth of a single target.

The proposed method can be applied to measure a depth of objects with structural surfaces. Furthermore, the dynamics of the measured objects is limited by the speed of the camera movement. However, it can be extended by using a variable opacity optical attenuation mask directly in front of a camera lens, [21].

For further research, the dithering approach can be applied with more steps in order to reduce the uncertainty to less than half.

Acknowledgements

The authors would like to thank Dr. Fredrik Bergholm at The Royal Institute of Technology, Sweden for his great contribution by lending their expertise regarding the necessary laboratory equipment and for his help. The authors would also like to acknowledge Dr. Siamak Khatibi at Blekinge Institute of Technology, Sweden for his continuous support and comments. Finally, we would like to thank Dr. Johan Höglund for his comments.

References

[1] W. Kulesza, J. Chen and S. Khatibi: "Arrangement of a Multi Stereo Visual Sensor System for a Human Activities Space". A. Bhatti (Ed.): *Stereo Vision*. InTech Education and Publishing, Vienna, 2008, pp. 153–172.

[2] J. Chen, S. Khatibi, W. Kulesza: "Planning of a Multi Stereo Visual Sensor System for a Human Activities Space". *Proc. of 2nd International Conference on Computer Vision Theory and Applications*, 2007, pp. 480–485.

[3] J. Chen, S. Khatibi, J. Wirandi, W. Kulesza: "Planning of a Multi Stereo Visual Sensor System for a Human Activities Space – Aspects of Iso-disparity Surface". *Proc. of SPIE on Optics and Photonics in Security and Defence*, vol. 6739, Florence, Italy, Sept., 2007.

[4] R. Hartley, A. Zisserman: *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[5] T. Chen, P. Catrysse, A. Gamal, B. Wandell: "How Small Should Pixel Size Be?". *Proc. of SPIE on Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, vol. 3965, 2000.

[6] H. Sahabi, A. Basu: "Analysis of Error in Depth Perception with Vergence and Spatially". *Computer Vision and Image Understanding*, vol. 63, no. 3, 1996, pp. 447–461.

[7] J. Chen, S. Khatibi, W. Kulesza: "Depth Reconstruction Uncertainty Analysis and Improvement – the Dithering Approach". *Elsevier Journal of Image and Vision Computing*, vol. 28, no. 9, 2010, pp. 1377–1385.

[8] C. Alasseur, A Constantinides, L. Husson: "Colour Quantisation Through Dithering Techniques". *IEEE International Conference on Image Processing*, vol. 1, 2003, pp. I-469-72.

[9] F. Corrêa Alegria: "Contribution of jitter to the error of amplitude estimation of a sinusoidal signal". *Metrol. Meas. Syst.*, vol. XVI, no. 3, 2009, pp. 465-478.

[10] *Two* new *perspective control wide-angle lenses*. Canon Technical Hall, *http://www.canon.com/cameramuseum/tech/report/200907/report.html.*

[11] A. Francisco, F. Bergholm: "On the Importance of Being Asymmetric in Stereopsis-or Why We Should Use Skewed Parallel Cameras". *International Journal of Computer Vision*, vol. 29, no. 3, 1998, pp. 181–202.

[12] M. Ben-Ezra, A. Zomet, S. K. Nayar: "Video Super-Resolution Using Controlled Subpixel Detector Shifts". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, 2005, pp. 977–987.

[13] A. Siddig: *Depth Reconstruction Uncertainty Improvement for Skewed Parallel Stereo Pair Cameras Using Dithering Approach*. Master Thesis, Dept. of Signal Processing, Blekinge Inst. of Tech., Karlskrona, Sweden, 2010. (submitted)

[14] *Image Processing Toolbox 7 User's Guide*. The MathWorks Inc., 1993, *http://www.mathworks.com/access/helpdesk/help/pdf_doc/images/image_tb.pdf*

[15] G. L. Mariottini, D. Prattichizzo: "EGT for Multiple View Geometry and Visual Servoing: Robotics Vision with Pinhole and Panoramic Cameras". *IEEE Robotics & Automation Magazine*, vol. 12, no. 4, 2005, pp. 26–39.

[16] *CCD Color Video Camera Module*. Sony Corp., 2002, h*ttp://*pro.*sony.com/-bbsccms/assets/files/mkt/indauto/Brochures/xc-555_techmanual.pdf*

[17] *Operating Manual MS 38E C-832 DC Motor Controller*. Physik Instrumente (PI) GmbH & Co., 1996, *http://www.physikinstrumente.net/ftpservice/Motor_Controllers/XXX__Oldcontrollers/C-832.DC-MotorController/C832.OperatingManual/MS38E280.pdf*

[18] *MegaPixel Camera*. Tamron Inc., *http://www.tamron.com/cctv/prod/23fm25sp.asp*

[19] W. Mustafa: *Depth Measurement Improvement Using Dithering Method in Sensor-shifted Stereo Cameras*. Master Thesis, Dept. of Signal Processing, Blekinge Inst. of Tech., Karlskrona, Sweden, 2010.

[20] J. Chen: "The Depth Reconstruction Accuracy in a Stereo Vision System" *XLI Intercollegiate Metrology Conference*, Gdansk, Poland, 2009.

[21] H. Farid: *Range Estimation by Optical Differentiation.* PhD Dissertation, University of Pennsylvania, Philadelphia, USA, 1997.

Paper IV

RFID Multi-Target Tracking Using the Probability Hypothesis Density Algorithm for a Health Care Application

Authors:

Jiandan Chen, Iyeyinka Damilola Olayanju, Olabode Paul Ojelabi and Wlodek Kulesza

Reformatted version of paper originally accepted to:

The 3rd International ICST Conference on IT Revolutions, Córdoba, Spain, March, 2011

RFID Multi-Target Tracking Using the Probability Hypothesis Density Algorithm for a Health Care Application

Jiandan Chen, Iyeyinka Damilola Olayanju, Olabode Paul Ojelabi and Wlodek Kulesza

Abstract

The intelligent multi-sensor system is a system for target detection, identification and information processing for human activities surveillance and ambient assisted living. This paper describes RFID multi-target tracking using the Gaussian Mixture Probability Hypothesis Density, GM-PHD, algorithm. The multi target tracking ability of the proposed solution is demonstrated in a simulation and real environment. A performance comparison of the Levenberg-Marquardt algorithm with and without the GM-PHD filter shows that the GM-PHD algorithm improves the accuracy of tracking and target position estimation significantly. This improvement is demonstrated by a simulation and by a physical experiment.

Keywords: *Human Tracking, Probability Hypothesis Density, Radio Frequency Identification.*

1. Introduction

The intelligent multi-sensor system, IMSS, is a high-performance autonomous distributed vision and information processing system, [1]. Fig. 1 illustrates the idea of using radio-frequency identification, RFID, sensors for person identification and localization. This is useful for medical healthcare services, security, the home help service etc., for all services that require robust tracking. One approach which has attracted particular attention is that of ambulatory wireless systems worn by elderly individuals in their own homes and designed to detect abnormalities in their motions and to report these abnormalities to a remote monitoring center for further action. The IMSS system inter alia includes vision, RFID sensors and actuators for surveillance and tracking of the human activities space. This space consists of the human beings and their surrounding environment, including robots, household appliances, and so on. An intelligent agent contains a knowledge database that includes learning and decision-making components that can be used to track, recognize, and analyze targets.

Just like the human eye, the multi vision sensor system acquires information about objects such as color, shape and position etc. in a 3D space. The accuracy of the target position estimation depends on many factors where the most important are: the distance between target and camera, camera focal length, stereo baseline and camera resolution, [2]. The target tracking robustness is also affected by light, obstacles, the camera's field of view etc. The human factor can also interfere with the tracking process. RFID is a rapidly developing technology based on wireless communication used to identify the target. The RFID system consists of tags which can be very small

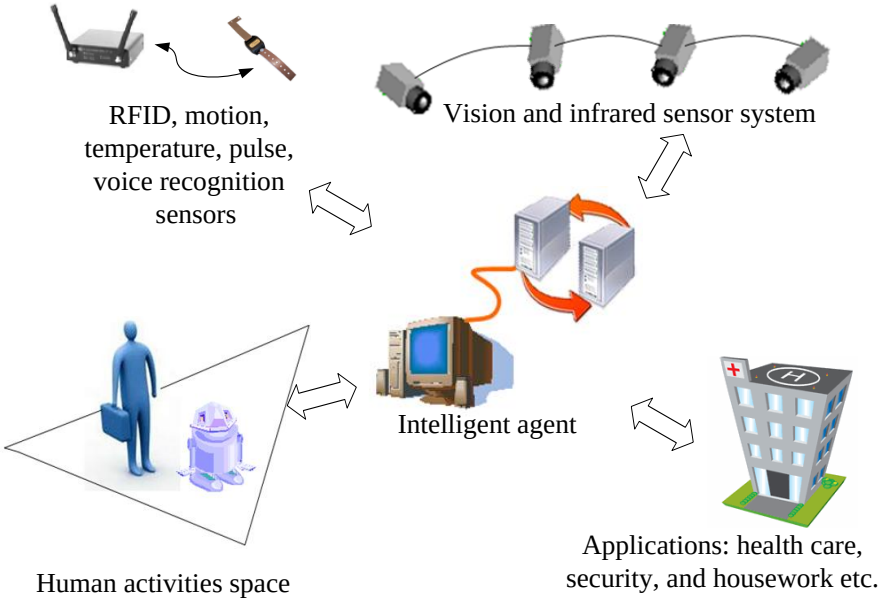


Fig. 1. Intelligent multi-sensor system

and easily carried by a human being. The tags can also be easily equipped with emergency buttons or other sensors, such as motion, temperature, heart rate, and voice recognition sensors. The RFID reader receives data from the tags that are used to identify the target. The technology has been widely studied and used in different applications, [3-5]. For instance, the tagged target's position can be tracked in an indoor environment by means of the tag's radio signal strength, RSS. The advantage of the RFID tracking system is its low-cost, large coverage area, independence of light, and ability to penetrate obstacles.

This paper proposes the use of the GM-PHD algorithm to track multi RFID targets. The estimated position of a RFID target can be then used by a vision sensor to identify the region of interest or the target's' initial position. Further, this information can be valuable when determining the tracking cameras' position, orientation, movement speed, focal length and baseline length, [1], [2]. In the case when the vision system is on standby due to occlusions, darkness or due to personal reasons, the RFID can still track the individual with a certain accuracy.

2. Related Works

The tracking of a specific target can be achieved by using different positioning systems like the Global Positioning System, GPS, ultrasonic, infrared and radio frequency signals. GPS can be used for an accurate outdoor localization while technologies like artificial vision, ultrasonic, infrared and radio frequency signals can be employed for indoor localization, [6-8]. The localization technique used in RFID relies mostly on an accurate estimation of the distance between the reader and the tag.

In a positioning system that uses ultrasonic or sonic signals, there is a need to initially deploy and distribute networked reference points within the tracking environment. The reference points form the reference space, [9]. Within the location where the system is used, special detectors are located to detect the tags' unique ultrasonic identification signals. As the monitored object is moving within the observed

area, the location of the object is being tracked and captured by the computer application (middleware), [10]. The position of the object is calculated with respect to the reference points. The integration of multi-sensor architecture for indoor navigation was introduced in [11].

The GM-PHD algorithm was proposed by Vo and Ma, [12]. Clark, Panta and Vo [13] suggested that the algorithm could be used for tracking multiple targets in a high clutter density. The comparison between the algorithm and the Multiple Hypothesis Tracker, MHT, shows that the former is better in areas of high clutter density and that it estimates the number of targets and their trajectories more accurately. Further, the performance of the GM-PHD algorithm for multi-target tracking in vision systems has been evaluated by Chen et al., [14].

The work of Sidenbladh discusses a particle PHD filter implementation for tracking multiple vehicles in terrain, [15]. The result shows the robustness of the method when tracking a varying number of targets.

3. Problem Statement and Main Contributions

The IMSS should instantly locate the multiple targets' initial positions when the system starts up. When the vision system is on standby or the targets are obscured by obstacles, the IMSS can still track the targets' motion by using the RFID system. The paper investigates the robustness of the RFID system when tracking multiple targets. The tracking improvement is accomplished by the application of the Levenberg-Marquardt, LM, algorithm with the GM-PHD filter.

The main contributions of the paper can be summarized as follows:

- Use of a minimization algorithm to estimate the propagation factor in the radio propagation channel model;
- Implementation of the GM-PHD filter on the LM algorithm for multi-target tracking using RFID technology;
- Verification of the GM-PHD algorithm for multi-target tracking in simulated and real environments.

4. Modelling

Radio propagation in an indoor environment is affected by reflections, diffractions and scatterings. The average radio signal power is decreased log-normally with the distance for indoor wireless channels. In this section, the radio propagation channel model is discussed. The estimation of the propagation factor using the minimum square error algorithm is proposed, and the LM algorithm with the GM-PHD filter is suggested to be used to track the multiple targets.

4.1 Radio propagation channel model

In wireless communication systems, the interaction between electromagnetic waves and the environment reduces the received signal strength. The path loss between two antennas strongly depends on the propagation environment. The power transfer ratio for a pair of lossless antennas in free space with optimum orientation is given by:

$$\frac{P_r}{P_t} = G_t G_r \left(\frac{\lambda}{4\pi d} \right)^2 \quad , \tag{1}$$

where λ is the wavelength; P_r is the received power; P_t is the transmitted power; G_r is the receiver antenna gain; G_t is the transmitter antenna gain; and d is the distance between antennas.

In equation (1) the factor $\left(\frac{\lambda}{4\pi d} \right)^2$, if separated from the effect of transmitter and receiver antenna gains, is referred to as the free space path loss. This is a path loss that occurs in a physical building and takes into account reflection, path obstruction, absorption and other attenuation effects introduced by the presence of objects inside the building [16].

The in-building path loss propagation model used to depict the effect of obstructions is given by:

$$PL(d) = PL(d_o) + 10\alpha_p \log\left(\frac{d}{d_o} \right) + AWGN \quad [dB] \quad , \tag{2}$$

where d_o is an arbitrary reference distance away; α_p is a propagation factor, i.e., the path loss exponent which depends on the surroundings and the building types; d is the transmitter-receiver separation distance; $PL(d_o)$ is the in-building path loss at an arbitrary reference distance away, which can be derived empirically; *AWGN* is additive white Gaussian noise with a zero-mean and the standard deviation σ.

The radio propagation properties can be estimated by the identification procedure. For instance, the optimized propagation factor, α_p, can be estimated by applying the minimum square error algorithm:

$$\min_{\alpha_p} \sum_{i=1}^{I} \left(PL_{RSS}(d_i) - PL(\alpha_p, d_i) \right)^2 \quad , \tag{3}$$

where PL_{RSS} is the measured path loss from RSS for the corresponding distance d_i from a reader to tag. The theoretical path loss, PL (α_p, d_i), is calculated from equation (2). I is the number of samples.

4.2 Target tracking using the GM-PHD algorithm

Using the wireless wave propagation model (2), we can estimate the distances from a tag to the multiple base stations, the RFID readers, using RSS with AWGN. For each reader as a center, a sphere can be formed with a radius equal to the distance from the station to the tag found from the RSS. The straightforward approach uses a geometric method to find the intersection point of the spheres around each reader. This point determines the tag position.

In practice, however, the calculated distances from a tag to the readers are contaminated by noise. The estimated spheres may not even have an intersection that allows the tag position to be found. In this case, one can apply the Levenberg-Marquardt, LM, algorithm to estimate the tag position, $\zeta(x, y, z)$, [17].

The distances, D_l, from a tag position, $\zeta(x, y, z)$, to the l-th reader's position, $\beta(x_l, y_l, z_l)$, can be described as:

$$D_l = \sqrt{(x - x_l)^2 + (y - y_l)^2 + (z - z_l)^2} \quad \forall\, l = 1, \ldots, L \quad , \tag{4}$$

where L is the number of readers.

The LM algorithm is an iterative technique that locates the minimum of a function expressed as:

$$\min_{\zeta(x,y,z)} \sum_{l=1}^{L} \left(D_l(\zeta) - d_l\right)^2 \quad , \tag{5}$$

where the distance from tag to readers d_l is calculated from (2) according to the RSS received by the l-th reader and where L is a number of readers.

In the framework of the multi-target tracker, the state of the multi-target position can be described by the Random Finite Set, RFS. The state can be represented as a discrete time k set $\mathbf{X}_k$ defined as:

$$\mathbf{X}_k = \left\{\chi_{k,i} : i = 1, \cdots, M_\chi(k)\right\} \quad , \tag{6}$$

where $M_\chi(k)$ is the number of targets in the space in the time k, and i is the index variable.

Firstly, according to (5) the state of the multi-target position can be estimated using the LM algorithm from the RSS measured by the RFID readers. Then, the multi-target measurement can be formulated as the set:

$$\mathbf{Z}_k = \left\{\zeta_{k,j} : j = 1, \cdots M_\zeta(k)\right\} \quad , \tag{7}$$

where $M_\zeta(k)$ is the number of observations in the time k, and j is the index variable**.**

The PHD algorithm assumes that each target generates observations independently and the clutter RFS is a Poisson distribution independent of target-oriented measurement, and that the predicted multi-target RFS is also a Poisson distribution. The GM-PHD filter is based on three additional assumptions compared to the PHD algorithm, [12]:

(i) Each target and the sensor follow the linear Gaussian model which can be described by:

$$\begin{aligned} f_{k|k-1}(\chi|\varsigma) &= N(\chi; \mathbf{F}_{k-1}\varsigma; \mathbf{Q}_{k-1}) \quad , \\ g_k(\zeta|\chi) &= N(\zeta; \mathbf{H}_k\chi; \mathbf{R}_k) \quad , \end{aligned} \tag{8}$$

where N is the normal or Gaussian distribution operator and $N(\cdot, m, \Psi)$ denotes a Gaussian density with the mean m and the covariance Ψ,
$\mathbf{F}_{k\text{-}1}$ is the state transition matrix,
$\mathbf{Q}_{k\text{-}1}$ is the process noise covariance matrix,
$\mathbf{H}_k$ is the observation matrix, and
$\mathbf{R}_k$ is the observation noise covariance matrix.

(ii) The survival and detection probabilities, p_S and p_D respectively, are state independent,

(iii) The intensity of the birth RFS is a Gaussian mixture.

Like other smoothing filters, the GM-PHD filter consists of two steps: prediction and update. The prediction equation is defined as:

$$v_{k|k-1}(\chi) = \gamma_k(\chi) + p_{S,k} \sum_{j=1}^{J_{k-1}} w_{k-1}^{(j)} N(\chi; m_{k-1}^{(j)}, \Psi_{k|k-1}^{(j)}) \quad , \tag{9}$$

where γ is the birth intensity and *w* is the weight parameter.

After the object detection has been finished and Z_k is available, the state is updated according to:

$$v_k(\chi) = (1 - p_{D,k}) v_{k|k-1}(\chi) + \sum_{\zeta \in Z_k} \sum_{j=1}^{J_{k|k-1}} w_k^{(j)}(\zeta) N(\chi; m_{k|k}^{(j)}, \Psi_{k|k}^{(j)}) \quad , \tag{10}$$

where *v* is the intensity function.

5. Implementation and validation

In this section, simulation results are shown to validate that the GM-PHD algorithm implemented in Matlab, can track multiple RFID targets. The performances of the LM algorithm with and without GM-PHD are analyzed and compared. Further, the measurements of the RFID system used in real environments verify the model and the simulation results. To assess the uncertainty of the estimated target position by the algorithm, we use the Euclidean distance between the actual location of the target and the estimated location.

5.1 Method validation by the simulation results

The models introduced in the previous chapter were implemented in Matlab and the validation scenario was then applied in a simulated space corresponding to a 3D indoor environment. This space was covered by signal receiving ranges of three RFID readers in order to determine the tag position. The simulation environment considered a regular room of the size 9 m × 6.5 m × 3 m as shown in Fig. 2. In order to simplify the analysis of the tracking process, the targets were represented as points in the 3D space and located at same height as the three readers. The readers were located at *A*(0, 0, 1.5) m, *B*(9, 0, 1.5) m, and *C*(9, 6.5, 1.5) m positions respectively. The propagation factor, α_p, was set as 3.00, which is consistent with the physical measurement shown the next section.

The model validation was carried out in a noisy environment. The RSS was simulated from the radio propagation path loss channel model (2) as a function of the distance between the tag and the reader, and contaminated by the AWGN. The noise takes input disturbances into account. In this simulation, the Gaussian noise distribution with the standard deviation 1.5 and the mean 0 was added.

To validate the GM-PHD algorithm implemented in Matlab and applied for RFID multi- target tracking, the trajectories of two targets in a form of circular motion and straight line motion respectively were used. The RFID tags' positions were first estimated by the LM algorithm (5) according to the distance from the tags to the different readers, where the distance was estimated using the path loss radio propagation channel model (2). Then, the GM-PHD algorithm was applied. In Fig. 2, the tag circular and straight line motion are illustrated as a red dash line and green solid line respectively, and the prediction positions from the GM-PHD algorithm are shown as red crosses and green asterisks respectively. Fig. 3 indicates the tracking manner of the

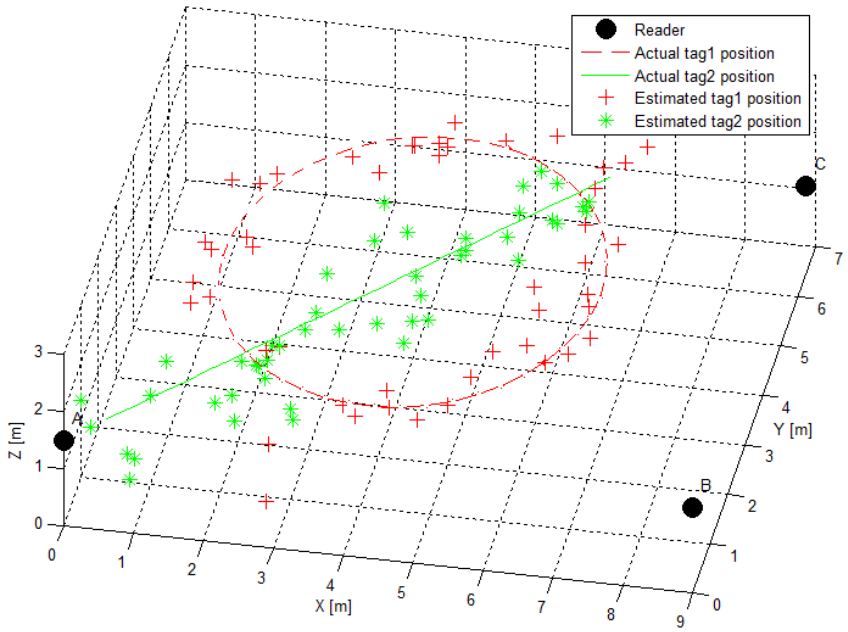


Fig. 2. The schematic diagram of the two targets motion paths in space simulation

GM-PHD algorithm in the *x* and *y* directions in respect to the time *k*. The same two truth trajectories as in Fig. 2 are presented as solid red and green lines, respectively. The GM-PHD predications of each of the two targets are marked as the red crosses and green circles respectively. The algorithm tracks the targets' movement successfully which confirms the GM-PHD algorithm's ability to track the multiple targets. The estimation uncertainties of the LM algorithm with and without GM-PHD are shown in Table I. Applying the GM-PHD algorithm reduces the mean and standard deviation of the distance estimation uncertainty by 30% and 33%, respectively.

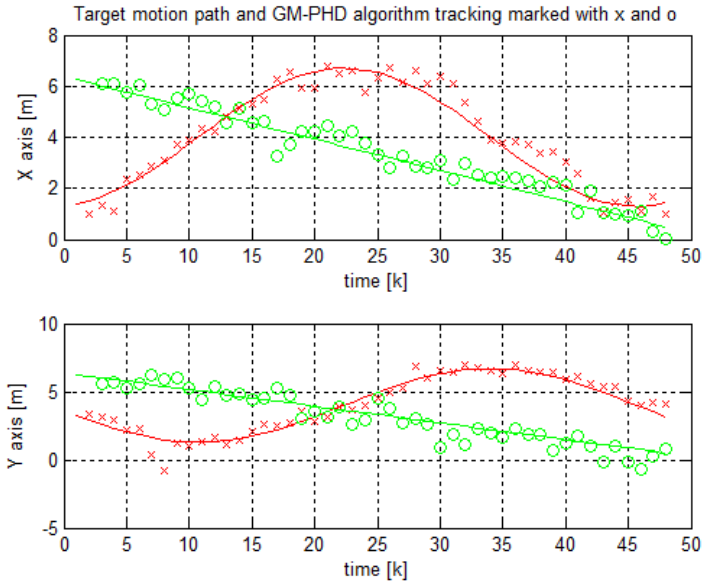


Fig. 3. The target motion paths in space from Fig. 2 in the x and y coordinates; for different target the ground truths are marked as the solid lines and the predictions by the GM-PHD are marked as crosses and circles respectively.

5.2 Physical Experiment

The active RFID system for target tracking consists of active tags and readers. The active tag used was a Wavetrend TG100-A, which is generally used for personnel tagging. The RFID reader was a Wavetrend RX-900 with an AN100 whip antenna, where the fronted side of the reader orients to the tag. The reader radio signal coverage was greater than 15 meters and at a working frequency of 433.92 MHz. The environment is considered as a free space of the size 9 m × 6.5 m × 3 m. The tag was mounted on a tripod and moved in the room without any obstacles and the tag antenna was kept in the vertical direction. The gross error was removed from the measurement of the radio signal strength, [18].

The propagation factor α_p was found using the optimization algorithm (3) for 20 samples when the tag moved away from the reader by a distance of 1 meter to 20 meters with an interval of 1 meter. Each sample was calculated as a mean value of three RSS measurements taken from each position of the tag. The propagation factor, α_p, was found to be 3.00.

Table I: The distance estimation uncertainty of the LM algorithm with and without the GM-PHD filter for the simulation and the physical experiment.

Experiment	Distance estimation uncertainty	With GM-PHD	Without GM-PHD
Simulation	Mean [m]	0.7	1.0
	Standard Deviation [m]	0.4	0.6
Physical	Mean [m]	1.0	1.4
	Standard Deviation [m]	0.6	0.9

Fig. 4 shows the schematic diagram of the experiment. Three readers were located at the $A(0, 0, 1.5)$ m, $B(9, 0, 1.5)$ m and $C(9, 6.5, 1.5)$ m positions respectively. Two tags were used in the real experiment and their trajectories are depicted in the figure as blue circles and turquoise squares respectively. The positions estimated by GM-PHD algorithm are marked as red crosses and green asterisks, respectively. Fig. 5 indicates the tracking manner of the GM-PHD algorithm in the x and y directions in the real environment in respect to the time k. The result verifies the GM-PHD algorithm ability to track a multi-target.

The estimation uncertainties when tracking real RFID tags using the LM algorithm with and without the GM-PHD filter are shown in Table I. Applying the GM-PHD algorithm reduces the mean and standard deviation of the distance estimation uncertainty with about 29% and 33% respectively, which validates the similar simulation results.

6. Conclusions

The paper proposes the LM algorithm with the GM-PHD filter to track multiple targets using the RFID system for ambient assisted living applications.

The propagation factor for the radio propagation model in the RFID system can be found by applying the optimal algorithm with the minimum square error. This robust and flexible method can be used to adjust the factor to different environments.

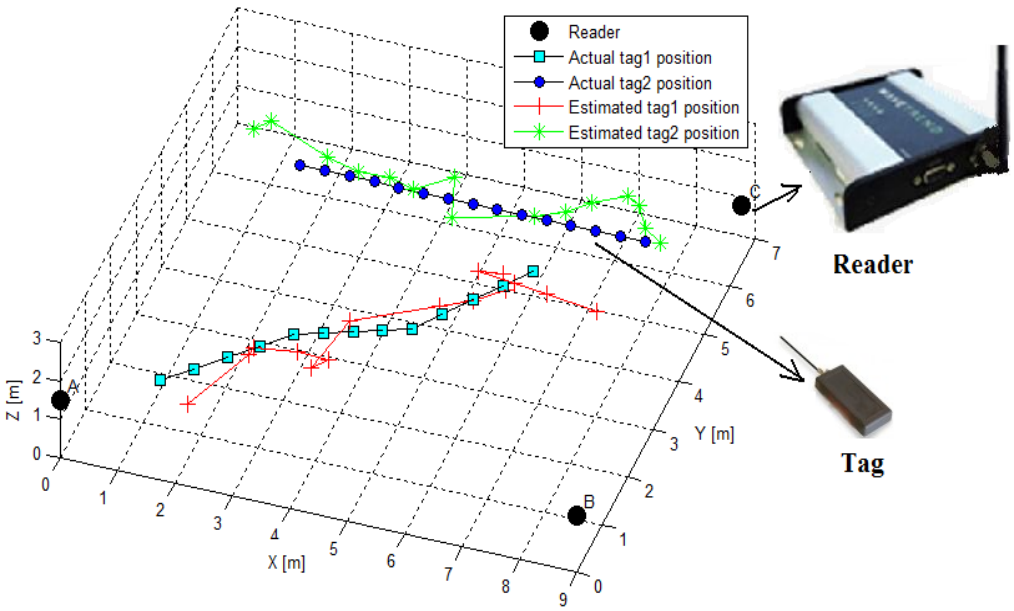


Fig. 4. The schematic diagram of the physical experiment

The LM algorithm with and without the GM-PHD filer has been implemented for multi-target tracking in RFID systems by Matlab. The ability of the GM-PHD algorithm for multi-target tracking is validated by simulation and real environment experiments. It is shown that the algorithm successfully tracks multiple targets while using the RSSs of each tag. The GM-PHD algorithm improves the accuracy of the target position estimation. The simulation and physical experiments show an improvement of the mean value by 30% and 29%, respectively, and an improvement of the both standard deviations by 33%, when compared to using the LM algorithm without the GM-PHD filter.

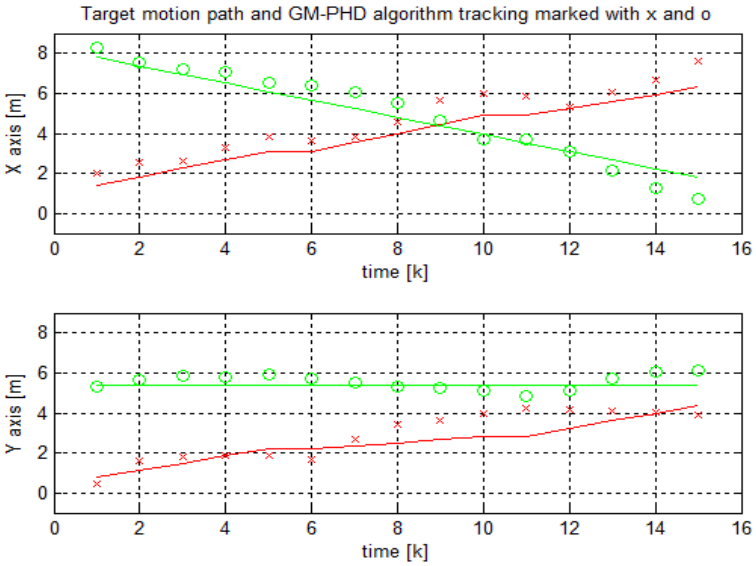


Fig. 5. The target motion paths in physical experiment environment in the x and y coordinates; for different targets the ground truths are marked as the red and green solid lines respectively and the predictions by the GM-PHD are marked as red crosses and green circles respectively.

In future work, the radio propagation model and the tracking algorithm should be investigated in more complex environments. The integration of the vision and RFID systems is going to be developed to use the tag position measured by the vision system as a reference to calibrate the RFID system. This lets the system automatically adjust the radio propagation factor according to changes in the environment.

Acknowledgements

The authors wish to acknowledge Dr. Anders M. Johansson for his inspiration and support, Dr. Johan Höglund for his comments and Mr. Wail Mustafa for help with the experimental data collection.

References

[1] Chen, J., Khatibi, S. and Kulesza, W., Planning of a Multi Stereo Visual Sensor System for A Human Activities Space, in: Proc. of the 2nd Int. Conf. on Computer Vision Theory and Applications, pp. 480-485 (2007)

[2] Kulesza, W., Chen, J. and Khatibi, S., Arrangement of a Multi Stereo Visual Sensor System for A Human Activities Space, in: A. Bhatti (Ed.), Stereo Vision, pp. 153-172, InTech Education and Publishing (2008)

[3] Yao, W., Chu, C. H. and Li, Z., The Use of RFID in Healthcare: Benefits and Barriers, IEEE International Conference on RFID-Technology and Applications (2010)

[4] Lin, C. J., Lee, T. L., Syu, S. L. and Chen, B. W., Applications of Intelligent Agent and RFID Technology for Indoor Position: Safety of Kindergarten as Example, in: Proceedings of the 9th International Conference on Machine Learning and Cybernetics (2010)

[5] Ouyang, D. F., Identification of Car Passengers with RFID for Automatic Crash Notification, Master Thesis, Blekinge Institute of Technology (2009)

[6] Abdelmoula, B., Horacio, S. and Mitsuji, M., RFID Indoor Positioning Based on Probabilistic RFID Map and Kalman Filtering, in: Proceedings of 3rd IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (2007)

[7] Bajaj, R., Ranaweera, S.L and Agrawal, D.P, GPS: Location Tracking Technology, E-Journal Computer, vol. 35, no. 4, pp. 92-94 (2002)

[8] Liu, H., Darabi, H., Banerjee, P. and Liu, J., Survery of Wireless Indoor Positioning Techniques and Systems, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 37, No. 6 (2007)

[9] Zhao, J. H. and Wang, Y. C., Autonomous Ultrasonic Indoor Tracking System, in: Proceedings of International Symposium on Parallel and Distributed Processing with Applications, pp 532-539 (2008)

[10] O'Connor, M. C., Testing Ultrasound to Track, Monitor Patients, RFID Journal, [Online] (2006), Available at: http://www.rfidjournal.com/article/articleprint/2199/-1/1/ [Accessed 13 Jan, 2010].

[11] Amanatiadis, A., Chrysostomou D., Koulouriotis, D., and Gasteratos A., A Fuzzy Multi-Sensor Architecture for Indoor Navigation, IEEE International Conference on Imaging Systems and Techniques, 452-457 (2010)

[12] Vo, B. and Ma, W. K., The Gaussian Mixture Probability Hypothesis Density Filter, IEEE Transactions Signal Processing, Vol. 54, No. 11, pp. 4091-4104 (2006)

[13] Clark, D., Panta, K. and Vo, B., The GM-PHD Filter Multiple Target Tracker, in: Proc. of 9th International Conference on Information Fusion, pp. 1-8 (2006)

[14] Chen, J., Adebomi O. E., Olusayo, O. S. and Kulesza, W., The Evaluation of the Gaussian Mixture Probability Hypothesis Density Approach for Multi-target Tracking, IEEE International Conference on Imaging Systems and Techniques, 182-185 (2010)

[15] Sidenbladh, H., Multi-Target Particle Filtering for the Probability Hypothesis Density, 6th International Conference on Information Fusion, pp. 800–806 (2003)

[16] Leong, K. S., Ling, M. and Cole, P. H., Positioning Analysis of Multiple Antennas in a Dense RFID Reader Environment, in: Proceedings of International Symposium on Applications and the Internet Workshops, pp. 56-59 (2006)

[17] Jorge, J. M., The Levenberg-Marquardt Algorithm: Implementation and Theory, in: Numerical Analysis, pp. 105-116 Vol. 630 (1978)

[18] McGhee, J., Henderson, I. A., Korczyński, M. J. and Kulesza, W.: Scientific Metrology, ISBN: 83-904299-9-3 (1998)

Paper V

Planning of a Multi Stereo Visual Sensor System for a Human Activities Space

Authors:

Jiandan Chen, Siamak Khatibi and Wlodek Kulesza

Reformatted version of paper originally published in:

Proceedings of the 2nd International Conference on Computer Vision Theory and Applications, pp. 480 – 485, Barcelona, Spain, March, 2007

Planning of a Multi Stereo Visual Sensor System for a Human Activities Space

Jiandan Chen, Siamak Khatibi and Wlodek Kulesza

Abstract

The paper presents a method for planning the position of multiple stereo sensors in an indoor environment. This is a component of an Intelligent Vision Agent System. We propose a new approach to optimize the multiple stereo visual sensor configurations in 3D space in order to get efficient visibility for surveillance, tracking and 3D reconstruction. The paper introduces a constraints method for modelling a Field of View in spherical coordinates, a tetrahedron model for target objects, and a stereo view constraint for the baseline of paired cameras. The constraints were analyzed and the minimum amount of stereo pairs necessary to cover the entire target space was optimized by an integer linear programming. The 3D simulations for human body and activities space coverage in Matlab illustrate the problem.

Keywords: *Sensor Placement, Multi Stereo View.*

1. Introduction

Vision is one of the most important information sources for humans. Human senses and the ability to process this information may be extended by the use of advanced technologies. The **I**ntelligent **V**ision **A**gent **S**ystem, IVAS, is such a high-performance autonomous distributed vision and information processing system. It consists of multiple sensors for gathering information and surveillance but also control of these sensors including their deployment and autonomous servo. It is able to extract 3D model information from a real scene of target objects, and compare this with a pattern in order to make decisions. Meanwhile the patterns are also renewed by the inclusion of a learning phase. These features enable the system to dynamically adjust camera configurations to track, recognize and analyze the objects, to achieve the desired 3D information. The Intelligent Agent consists of a knowledge database, with learning and decision making components. Figure 1 shows the block diagram working sequence of the IVAS. The paper focuses on the planning of stereo pair deployment of the system.

The critical problem for the system is to find the optimal configuration of sensors so that the features of the environment and target objects are visible under the required constraints. The sensors' intrinsic and extrinsic parameters are examples of parameters considered while choosing the configuration. The system also requires optimal configuration for stereo pair design.

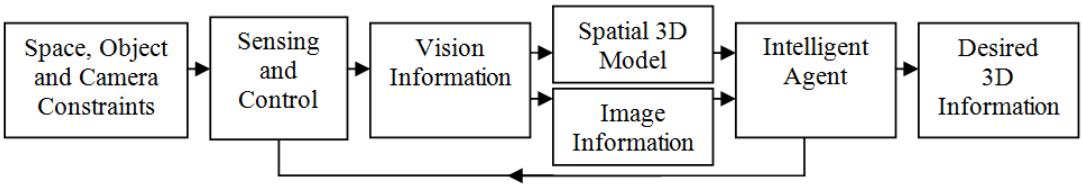


Figure 1: The block diagram of Intelligent Vision Agent System.

1.1 Related works

The sensor planning can be viewed as an extension to the well-known **A**rt **G**allery **P**roblem, AGP, (O'Rourke, 1987). In its simplest form, the AGP describes a simple polygon, often with holes, and the task is to calculate the minimum number of guards necessary to cover the entire polygon. Sensor planning has a similar goal, to minimize the number of sensors needed to cover the target space. The AGP has a **F**ield **o**f **V**iew, FoV, of 360° for the guards, whereas in sensor planning the camera's FoV is limited by image resolution and the viewable angles of cameras. A stereo view requires the target space to be covered by at least two views.

Camera placement algorithms based on binary optimization techniques are known, and analyzed for camera deployment in a polygonal space in 2D space (Erdem&Sclaroff, 2006). Also a linear programming method to optimize sensor placement, with respect to coverage, has been developed (Hörster&Lienhart, 2006; Chakrabarty et al., 2002). Using the sensor detection range *r* to solve the area of grid coverage problem is common (Hörster &Lienhart, 2006; Chakrabarty et al., 2002; Zou &Chakrabarty, 2004). A quality metric including a probabilistic occlusion can be used to evaluate the optimum configurations of multiple cameras (Chen, 2002). The sensor planning can be analyzed by examining the visibility in the dynamical environment, and the result simulated by re-annealing software (Mittal, 2006). An optimal stereo vision configuration for a mobile robot focuses on optimizing the stereo pair orientation to detect static obstacles where the stereo pair is assumed to be known a priori (Huang&Krotkov, 1997). For a model-based sensor placement, the target geometry information is known (Fleishman et al., 2000; Chen&Li, 2004). Mobile single camera positioning to optimize the observability of human activity has been studied (Bodor et al., 2005). There has been relatively little work on determining optimal multiple sensors for sensor configurations (Mittal, 2006).

2. Problem formulation

The algorithm proposed in the paper works in 3D space and a new approach to define the camera's FoV, applied in spherical coordinates, is proposed. In the presented solution the maximum volume of FoV coverage becomes a part of sphere that simplifies the calculation. The definition is intrinsically related to the sensor's physical parameters, such as the dimension of the CCD and focus length. For the camera's view, this paper considers not only the problem of coverage, but also the orientation of the target. To deal with this, a target space is modelled by a tetrahedron. The presented method formulates all factors into the constraints, and has a flexible way to add other constraints. Knowledge of stereo technology is integrated, a greedy stereo pair search algorithm solving for the minimal amount of stereo pairs by **I**nteger **L**inear **P**rogramming, ILP, is proposed and the ILP model is given.

2.1 Problem statement and main contributions

The paper addresses the problem of determining the optimum amount of cameras and corresponding positions and poses to observe human body and activities space in stereo views.

The main contributions of the paper may be summarised as follow:

- The new approach to modelling a 3D FoV using spherical coordinates;
- Modelling of human and target space as tetrahedrons;
- Stereo pairs formulation by a greedy algorithm using stereo constraints;
- Minimizing the amount of stereo pairs by means of the stereo view integer linear programming model.

2.2 Definitions and constraints

The space denotes a 3D indoor environment. The target object or space describes the space for human body and activities, and is required to be covered by cameras' FoVs. In other words, it should be visible to the cameras and respect the minimal requirements of each constraint. The constraints analysis ensures sufficient data of scene features for 3D reconstruction and image analysis. Design of the optimal parameters for cameras' positions, poses and stereo baseline length is done according to the criteria from cameras' FoVs; the target objects and stereo matching.

The following factors formulate the constraints:

Field of View is the maximal space volume visible from a camera. The FoV is a cone determined by the azimuth and elevation within a spherical coordinate system.

***I**mage **R**esolution,* IR, describes the visibility of the object in a camera view as the size of the object in the image. IR is affected by the distance from camera to the target object and the angle between the camera view direction and the orientation of the target objects surface.

Stereo Baseline Length is the distance between the paired cameras in a stereo view**.** Stereo matching becomes harder when the baseline length increases.

2.2.1 Camera constraints

The horizontal and vertical viewable angles of the camera can be determined by the focal length of the lens and the size of the CCD element:

$$\phi_h = 2\arctan\frac{S_h}{2f},\ \phi_v = 2\arctan\frac{S_v}{2f}, \tag{1}$$

where ϕ_h and ϕ_v define the horizontal and vertical viewable angles of the camera FoV; S_h, S_v are the horizontal and vertical dimensions of the CCD element, and f is the focal length of the lens.

The camera working distance, r, is the radius of a sphere and can be calculated from the focal length of the lens f and image resolution requirement.

The camera position $C(x_c,y_c,z_c)$ and pose $\psi(\alpha_c,\beta_c)$ describe the camera's extrinsic parameters. The camera pose defines its azimuth α_c and elevation β_c.

In the world frame, the target object and camera's position and pose are described in Cartesian coordinates. In the camera view, a spherical coordinate system is applied. The distance l between the target position $O(x,y,z)$ and camera position $C(x_c,y_c,z_c)$ is:

$$l = \sqrt{(x-x_c)^2 + (y-y_c)^2 + (z-z_c)^2} \ . \tag{2}$$

The azimuth α_o and elevation β_o of target object with respect to camera position are given by

$$\alpha_o = \arctan\frac{x-x_c}{y-y_c}, \ \beta_o = \arcsin\frac{z-z_c}{l} \ . \tag{3}$$

In order for the target object feature point to be covered by the camera's FoV, the following constraints must be fulfilled:

$$\begin{aligned} l \le r \ \text{ and} \\ \alpha_c - \phi_h/2 \le \alpha_o \le \alpha_c + \phi_h/2 \ , \\ \beta_c - \phi_v/2 \le \beta_o \le \beta_c + \phi_v/2 \ . \end{aligned} \tag{4}$$

In the spherical coordinate systems, the range of the camera's FoV is directly determined by S_h, S_v, r and f, which makes it easy to dynamically compute FoV according to the changing of the focal length f. The modelling of the FoV can be viewed as a part of the sphere, as shown in Figure 2.

2.2.2 Object constraints

In the human living environment, we always have some knowledge about the target objects and space under observation, e.g. the floor plan of the room, the geometric properties of the furniture, human body and activities space, etc.

The 3D target object or space can be modelled by a tetrahedron, giving four triangles. We define four vertices of tetrahedron by $Tv_{1,2,3,4}$, as in Figure 3. The three upward triangles are required to be covered by cameras' FoVs. The normal of each

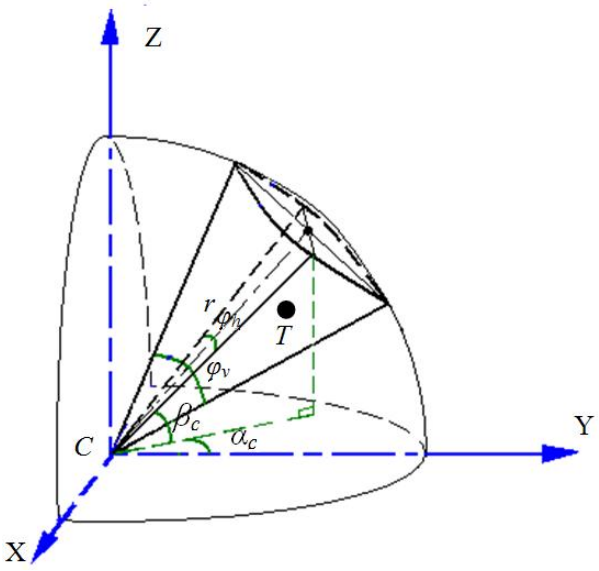


Figure 2: The spherical co-ordinates system and FoV of a camera where C is camera position and the example target point is located at point T.

triangle gives the orientation of the surface. If the visibility angle θ, between the triangle normal $\vec{n}$ and a line drawn from the centroid of triangle to camera position increases then the image resolution decreases. In order to get good image resolution, an angle θ less than the maximum visibility angle θ_{max} is required:

$$\theta \leq \theta_{\max} . \tag{5}$$

It is best that the camera orientation $\vec{c}$ lines up with the centroid of triangle, bringing the target object to the centre of the camera's FoV and causing less lens distortion. The angle between camera orientation $\vec{c}$ and a line drawn from camera position to the centroid of triangle less than the maximum φ_{max} is also required and constrained as:

$$\varphi \leq \varphi_{\max} . \tag{6}$$

The triangle is considered to be covered if all three vertices are within a camera's FoV and fulfil constraints (5) and (6), guaranteeing good observability of the target object.

2.2.3 Stereo pair constraints

We construct the stereo coverage from the overlap of two cameras' FoVs. Overlapping FoVs are typically used in computer vision for the purpose of extracting 3D information (Khan et al., 2001). The area of stereo coverage must cover all of the target objects. Assuming the camera is a pinhole camera, the 3D depth *Z* is given by (Faugeras, 1993):

$$Z = \frac{Bf}{dx}, \tag{7}$$

where *B* is the baseline length between two cameras and dx is the disparity.

The accuracy of depth resolution relies on stereo matching, but stereo matching becomes harder as the baseline length increases. Hence, we have a constraint defining

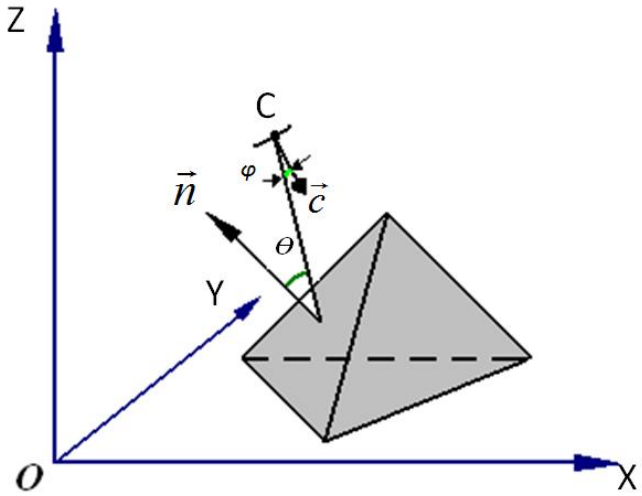


Figure 3: Illustration of the human space modelled as a tetrahedron; θ - the visibility angle between the triangle normal $\vec{c}$ and a line from the centroid of the triangle to the camera position; φ - the angle between the camera orientation $\vec{c}$ and a line from the camera position to the centroid of the triangle.

the maximum baseline length for stereo matching:

$$B \le B_{\max} . \tag{8}$$

3. Approaches

The stereo pair placement problem consists of two stages. Firstly, we find potential stereo pairs that satisfy stereo constraint by greedy searching from all potential cameras' positions and poses. Secondly, we minimize the amount of stereo pairs needed, subject to the coverage constraint.

3.1 Greedy algorithm

The algorithm gives a flexible way to organize cameras into stereo pairs, each potential camera to be included in a stereo pair may be chosen by an algorithm according to the stereo pair constraint. The first step of the algorithm is to sample the potential camera's positions $C_n(x_{cn},y_{cn},z_{cn})$ and poses $\psi_n(\alpha_{cn},\beta_{cn})$ of the camera state, $Scamera^{k}_{C_n,\psi_n}$, where k is camera state index number. The target object, which we must cover, is modelled as a tetrahedron. In the next step, we compute all of the potential cameras' positions and poses needed to cover each upward triangle of this model. Taking this, we combine every two camera states to be a potential stereo pair, $Stereopair_i$, according to the stereo constraint (8). The algorithm is sufficiently flexible to add other constraints for stereo pairs, e.g. the angle constraint between the cameras' optical axes. Finally the algorithm removes the redundant potential stereo pairs.

3.2 Stereo view integer linear programming model

This model assumes that one type of camera is used throughout, resulting in just one camera's FoV being considered. The optimization of the amount of cameras with different FoVs also can be easily extended, by adding one more term for different FoVs. Since the stereo pairs have been found by the greedy algorithm, the integer linear programming can be applied to minimize the total stereo pairs subject to the coverage constraint (Hörster &Lienhart, 2006; Chakrabarty et al., 2002).

A binary variable is computed and stored in advance. The stereo visibility binary variable table $Stereovis_{j,i}$ is defined by:

$$Stereovis_{j,i} = \begin{cases} 1 & \text{if a } Stereopair_i \text{ covers} \\ & \text{triangle } j \text{ of target object model,} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

which indicates each triangle j as the row j to be covered by the stereo pair i in the column i, and $1 \le i \le K_s$, where K_s is the total number of stereo pairs.

This objective function minimizes the number of stereo pairs needed to cover all triangles in the target object model, and also ensures that the target object is covered by at least one stereo pair:

$$\min \sum_{i=1}^{K_s} S_i , \tag{10}$$

subject to

$$\sum_{i=1}^{K_s} S_i \times Stereovis_{j,i} \geq 1, \quad \text{for } j = 1,2,3, \tag{11}$$

where the S_i is the binary variable where a "1" indicates the stereo pair to be chosen.

To ensure that only one camera is located at each position and has only one pose, the conflict binary variable table $c_{p,i}$ is also calculated in advance and defined by:

$$c_{p,i} = \begin{cases} 1 & \text{if two pairs } i \text{ and } p \text{ share the} \\ & \text{same camera with different} \\ & \text{orientations, where } i \neq p \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

for i=1, 2, …, K_s, and p=1, 2, …, K_s.

One more constraint is added into the model:

$$\sum_{i=1}^{K_s} S_i \times c_{p,i} \leq 1, \quad \text{for } p = 1,2,\cdots,K_s. \tag{13}$$

The information on the optimal number of stereo pairs, and which pairs to use, are returned as vectors by the ILP model.

4. Results

The described algorithm was simulated in MATLAB 7.0. The integer linear programs *lpsove package* (Berkelaar et al., 2005) and the Epipolar Geometry Toolbox (Mariottini&Prattichizzo, 2005) were used to minimise the amount of cameras and transform the object position in 3D separately. The simulation environment considers a rectangular room with size 8x8x3 m. The modelling of the human body as a tetrahedron requires three upward triangles; each of them occludes the triangles behind it and must be visible to at least one pair of cameras. The human model is 2 m high and 1.2 m at the base edges. The cameras' positions are restricted to the ceiling around the room, their potential positions sampled at half meter intervals, and the poses sampled at 12° intervals. The camera has same horizontal and vertical viewable angles ϕ_h, ϕ_v of 60° and has a working distance r of 7 m. The maximum visibility angle θ_{max} and the angle φ_{max} are taken to be 70° and 10° respectively. The maximum stereo baseline length B_{max} is 1.5 m.

This case study illustrates the optimum amount of stereo pairs with corresponding cameras' positions and poses changing according to the model location. In order to clearly show cameras' positions and poses, the analysis only considers the model at three locations 1, 2 and 3, see Figure 4. The arrows indicate the optical axes of the cameras. The index numbers indicate the model locations and corresponding cameras' positions and poses. In each position every upward triangle surface is visible to at least one stereo pair; the algorithm proves that a set of two pairs is sufficient to cover three triangle surfaces. When the model moves from position 1 to position 2, the stereo pair

positions (0,200) and (0,250) change to (0,0) and (0,50) respectively. The elevation angle is increased as the model moves further away from the camera. At the same time, another stereo pair located at (600,0) and (650,0) moves to (800,100) and (800,150) respectively, the elevation angle is decreased as the model moves closer to it. The azimuth α_c and elevation β_c in stereo pair may vary by camera individually. Both two stereo pairs follow the model when the model changes from position 2 to position 3, see Figure 4.

5. Conclusion

The proposed approach is useful in determining the optimal number of cameras and their corresponding positions and poses to observe human body and activities space in stereo view. The stereo pair has the flexibility to adjust cameras' poses and positions individually. Multi camera planning and control for surveillance and tracking in supermarkets, museums and the home environment, and especially in situations which require stereo data to reconstruct 3D, are possible fields of application.

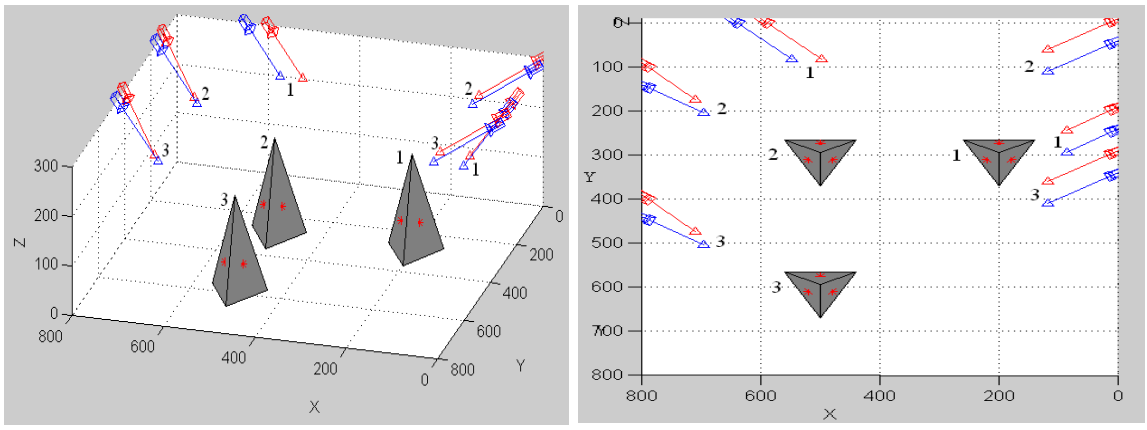


Figure 4: The human space modelled as tetrahedron with corresponding cameras' positions and poses changing according to the model location; (a) perspective view (b) top view (c) side view.

To model the target object as a tetrahedron gives a convenient way to extract the orientation of each surface and guarantee a good observability. Modelling camera's FoV using spherical coordinates simplifies the model and constraints, which speeds up computations. Formulating the stereo pairs with greedy algorithm using stereo constraints is a simple way to get all possible stereo pairs and then minimize the amount of stereo pairs by means of the stereo view ILP model.

It is possible to extend this algorithm to dynamic cameras to track humans. In order to follow target objects movement, the camera movement distance constraints can be applied (Chen et al., 2007). The human activities space also can be extended to a large space modelled by multiple tetrahedrons. The space can be covered without changes of cameras' positions and poses. Future work may focus on dynamic occlusions and tracking multiple dynamic objects by using multiple dynamic stereo pairs.

References

Berkelaar, M., Notebaert, P., and Eikland, K., 2005, Lpsolve 5.5: Open Source (mixed-integer) Linear Programming System. Eindhoven Univ. of Technology, http://tech.groups.yahoo.com/group/lp_solve/files/.

Bodor, R., Drenner, A., Janssen, M., Schrater, P., Papanikolopoulos, N., 2005, Mobile Camera Positioning to Optimize the Observability of Human activity Recognition Task, IEEE/RSJ Int. Conf. on *Intelligent Robots and Systems.*

Chakrabarty, K., Iyengar, S. S. , Qi, H., and Cho., E., 2002, Grid Coverage for Surveillance and Target Location in Distributed Sensor Networks. *IEEE Transaction on Computers*, 51(12): 1448-1453.

Chen, J., Khatibi, S., Kulesza, W., 2007, Planning of A Multi Stereo Visual Sensor System Depth Accuracy and Variable Baseline Approach. *3DTV-Con, the True Vision Capture, Transmission and Display of 3D Video*.

Chen, S. Y., Li Y. F., 2004, Automatic Sensor Placement for Model-Based Robot Vision. *IEEE Transactions on Systems, Man, and Cybernetics*, 34(1):393-408.

Chen, X., 2002, Design of Many-Camera Tracking Systems for Scalability and Efficient Resource Allocation. PhD thesis, Stanford University.

Erdem, U., Sclaroff, S., 2006, Automated Camera Layout to Satisfy Task-Specific and Floor Plan-Specific Coverage Requirements. *Computer Vision and Image Understanding*, 103(3): 156-169.

Faugeras, O., 1993, *Three-dimensional computer vision*. MIT Press.

Fleishman, S., Cohen-Or, D., and Lischinski, D., 2000, Automatic Camera Placement for Image-based modelling. *Computer Graphics Forum*, 19(2):101-110.

Huang, W. H. , Krotkov, E. P., 1997, *Optimal Stereo Mast Configuration for Mobile Robots.* Processing of IEEE Int. Conf. Robotics.

Hörster, E., Lienhart, R., 2006, On the Optimal Placement of Multiple Visual Sensors, *ACM International Workshop on Video Surveillance & Sensor Networks*.

Khan, S., Javed, O., Rasheed, Z., Shah, M., 2001, Human Tracking in Multiple Cameras. *The Eighth IEEE Int. Conf. on Computer Vision*.

Mariottini, G. L., Prattichizzo, D., 2005, The Epipolar Geometry Toolbox: Multiple View Geometry and Visual Servoing for Matlab. *Proceedings of the IEEE Int. Conf. on Robotics and Automation*.

Mittal, A., 2006, Generalized Multi-Sensor Planning. *9th European Conference on Computer Vision (ECCV).*

O'Rourke, J.,1987 *Art Gallery Theorems and Algorithms*. Oxford University Press.

Zou, Y. and Chakrabarty, K., 2004, Sensor Deployment and Target Localization in Distributed Sensor Network. *Trans. on Embedded Computing Sys*., 3(1):61-91.

Paper VI

Planning of a Multi Stereo Visual Sensor System - Depth Accuracy and Variable Baseline Approach

Authors:

Jiandan Chen, Siamak Khatibi and Wlodek Kulesza

Reformatted version of paper originally published in:

Proceedings of IEEE Computer Society 3DTV-Con, the True Vision Capture, Transmission and Display of 3D Video, Kos, Greece, May, 2007

Planning of a Multi Stereo Visual Sensor System - Depth Accuracy and Variable Baseline Approach

Jiandan Chen, Siamak Khatibi and Wlodek Kulesza

Abstract

This paper presents a method for planning the position of multiple stereo sensors in an indoor environment. This is a component in an Intelligent Vision Agent System. We propose a new approach to dynamically adjust the multiple stereo pair's position, pose and baseline length in 3D space in order to get sufficient visibility and enough accuracy for surveillance, tracking and 3D reconstruction. The paper proposes visibility constraints to plan the camera's pose, and a depth accuracy constraint to control the baseline length. The minimum number of stereo pairs necessary to cover the target space is optimized by an integer linear programming. The 3D simulations of reconstruction accuracy and the human activities space coverage problem were performed in Matlab.

Keywords: *Active Vision, Image Reconstruction, Position Control, Stereo Vision.*

1. Introduction

Vision is one of the most important information sources for humans. Human senses and abilities to process this information may be extended by the use of advanced technologies. The Intelligent Vision Agent System, IVAS, is one such high-performance autonomous distributed vision and information processing system. The IVAS has learning and decision making functions. For such a system, the critical problem is to find the optimal configurations of sensors and to control a reconstruction accuracy level, so that the features of environment and the target object are visible under the required constraints. The stereo pair's position, pose and baseline length are the most important factors for the 3D reconstruction. The agent system can apply a depth accuracy constraint according to the requirements to get sufficient accuracy, but too high accuracy requirements will restrict a camera movement unduly.

Sensor placement algorithms based on binary optimization techniques are known, and have been analyzed for camera deployment in 2D space, e.g. a linear programming method to optimize sensor placement with respect to coverage has been developed, [1], [2]. The optimum number of cameras and corresponding positions and poses to observe human activities space by a linear programming model is known, [3]. The planning of the sensors can be carried out by examining visibility in the dynamical environment, and the result can be simulated by a re-annealing software, [4]. In the field of active vision, there have been some studies on the dynamical adjustment of stereo

baseline for one stereo pair to improve the reconstruction accuracy, [5], [6]. However, there has been relatively little work on determining optimum sensor configurations, [4].

2. Problem formulation

The paper addresses the problem of determining the multiple stereo pair's baseline lengths, corresponding positions and poses to observe the human activities space with a depth accuracy requirement.

2.1 Problem statement and main contributions

The algorithm proposed in the paper works in 3D space. The human activity space is defined by a tetrahedron. The approach dynamically adjusts the stereo pair's baseline length according to the accuracy requirement and the target distance as a distance from the target position to the stereo pair baseline. The 3D reconstruction accuracy, which is ensured by the accuracy constraint, can be verified by a cubic reconstruction. The minimum quantity of stereo pairs to cover the human activity space is solved by means of Integer Linear Programming, ILP.

The main contributions of the paper may be summarised as follow:

- To analyse the key factors which affect the accuracy of 3D reconstruction, assuming that a depth quantization error is given by a single image pixel;
- To control the stereo pair's baseline length, position and pose by means of a depth accuracy constraint, which guarantees a certain accuracy in the 3D reconstruction;
- To implement a cubic reconstruction, to verify the reconstruction accuracy;
- To apply the two stages sampling rate and the limited movement range for the camera positioning system, where a new camera position is found based on the previous position and a movement distance constraint.

2.2 Definitions and constraints

The space denotes a 3D indoor environment and the target is a human activity space covered by the camera's **F**ield **o**f **V**iew, FoV. We implemented three types of constraints to ensure sufficient data from the 3D scene.

2.2.1 Visibility and baseline length constraints, [3]

The camera constraints describe the camera's FoV. In the spherical coordinate systems, the camera's horizontal and vertical viewable angles, ϕ_h, ϕ_v, and a working distance, r, can be calculated from camera's attributes.

In order for the target object's feature points to be covered by the camera's FoV, the following constraints must be fulfilled:

$$\begin{gathered} l \le r \text{ and} \\ \alpha_c - \phi_h / 2 \le \alpha_o \le \alpha_c + \phi_h / 2, \\ \beta_c - \phi_v / 2 \le \beta_o \le \beta_c + \phi_v / 2, \end{gathered} \tag{1}$$

where: l is the distance between the target position and camera's position; α_o, β_o are respectively the azimuth and elevation of target; α_c, β_c are respectively the azimuth and elevation of the camera's pose.

The human activities space is modelled by a tetrahedron. The normal of each triangle gives the orientation of its surface. If the visibility angle, θ, between the triangle normal and a line drawn from the centroid of triangle to the camera's position, increases then the image resolution decreases. In order to get a good image resolution, the visibility angle, θ, of less than the maximum visibility angle, θ_{max}, is required:

$$\theta \le \theta_{\max}. \tag{2}$$

The camera's orientation should line up with the centroid of triangle, thus bringing the target object to the centre of camera's FoV and causing less lens distortion. The angle between the camera's orientation and the line drawn from camera's position to the centroid of triangle, φ, of less than the maximum φ_{max} is required and is constrained as:

$$\varphi \le \varphi_{\max}. \tag{3}$$

Since stereo matching becomes more difficult when the baseline distance increases, the baseline length has to be limited to the maximum stereo baseline length, B_{max}:

$$B \le B_{\max}. \tag{4}$$

2.2.2 Depth accuracy constraints

We construct the stereo coverage from the overlapping of two cameras' FoVs. The overlapping FoVs are typically used to extract 3D information. The area of stereo coverage must cover all the target objects. In the most common case, the cameras form a converging stereo pair. Cameras' poses azimuths and baseline are shown in Fig. 1. Cameras' convergence angles, (α_{cl}, α_{cr}), are the angles of each camera rotated inwards from the parallel to achieve convergence. The target convergence angles, (ψ_l, ψ_r), are the angles between the visual lines of each camera and the baseline perpendicular. From Fig. 1, simplifying: $\psi = \psi_l = \psi_r$ and $\alpha_c = \alpha_{cl} = \alpha_{cr}$, we obtain:

$$Z = \frac{B}{2\tan\psi}, \tag{5}$$

where: B is a baseline length and Z is a target distance.

The equation (5) can be written as:

$$Z = \frac{B}{2}\cdot\left(\frac{1}{\tan\alpha_c - \tan(\alpha_c - \psi)} + \frac{\tan\alpha_c \cdot \tan(\alpha_c - \psi)}{\tan\alpha_c - \tan(\alpha_c - \psi)}\right). \tag{6}$$

In the case of parallel stereo or with the target close to the fixation point, the α_c or (α_c-ψ) varies by a small amount, and the equation (6) can be further simplified. The resolution of the target convergence angle, ψ, is related to a single pixel, p, in the image, thus the relative depth error can be written as:

$$\frac{\Delta Z}{Z} \approx \left|\frac{\cos\alpha_c}{\sin\psi\cos(\alpha_c - \psi)}\right| \cdot \frac{p}{f} = AF \cdot \frac{p}{f}, \tag{7}$$

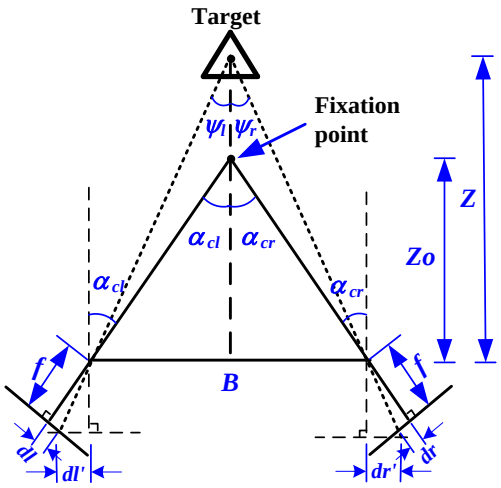


Fig. 1. The model of stereo pair geometry; Z_0 is the depth of the fixation point, $|dl\text{-}dr|$ is stereo disparity, f is focal length.

where: *AF* is the depth accuracy factor, *f* is a focal length, and the depth quantization error is assumed to be one pixel, *p*.

The depth error is proportional to the depth accuracy factor. In fact, since the depth accuracy factor varies more significantly with respect to the target convergence angle, ψ, than to the camera's pose, α_c, therefore the target convergence angle determines the depth accuracy factor. The accuracy constraint for a given *p* can be defined as:

$$AF \le AF_{con}\,, \tag{8}$$

where: AF_{con} is determined from the reconstruction accuracy requirements of the given application.

2.2.3. Camera movement distance constraints

In order to follow the movement of target object, camera's movement distance constraint can be applied. The next-view position for the camera should not be placed too far away from the previous one. This constraint is formulated as:

$$Dist(StereoPair_{next}, StereoPair_{current}) \le Dis_{\max}\,, \tag{9}$$

where: Dis_{max} is the camera's maximum movement distance.

This constraint for obvious reasons will simplify computation.

3. Approach

Solving the stereo pair placement problem utilises the three distinct stages.

- Firstly, the depth accuracy factor constraint, AF_{con}, is applied for the target. All potential stereo pairs that satisfy the constraint are found by a greedy algorithm, [3].

- Secondly, to minimize the amount of stereo pairs needed, the optimum number of stereo pairs is subject to the coverage constraint by the stereo view integer linear programming model, [3].
- Finally, in order to verify 3D reconstruction accuracy, the cubic reconstruction is simulated using a pair of rectified scene images. The rectification matrix is computed directly from the **p**erspective **p**rojection **m**atrix, PPM, [7], and the rectification algorithm also gives two new PPMs, P_n1 and P_n2. The cubic reconstruction in 3D can be performed with a triangulation method directly from the rectified images, using P_n1, P_n2. The 3D reconstruction error, Δ_{rec}, for a single pixel error along a horizontal direction in the rectified image, has same value as the depth error and is given by:

$$\Delta rec = \sqrt{\frac{1}{8}\sum_{i=1}^{8}\left|\hat{M}_i - M_i\right|^2}\ , \tag{10}$$

where $\hat{M}_i$ gives the coordinates of the reconstruction point i in a cube of the rectified images and M_i gives the real coordinates of the target point i in the cube.

4. Results

The simulations were performed in MATLAB 7.0. The integer linear programs *lpsove package*, [8], and the Epipolar Geometry Toolbox, [9], were used to minimise the number of cameras and transform the object position in 3D separately. The simulation environment considers a rectangular room with size 8 m × 8 m × 3 m. The modelling of the human activities space as a tetrahedron requires three upward triangles; and each triangle must be visible to at least one pair of cameras. Each model is 2 m high and 1.2 m at the base edges. The cameras' positions are restricted to the ceiling around the room, their potential positions sampled at 0.2 m intervals in the initial phase, and 0.1 m intervals for a next view camera's position; Dis_{max} is taken 3 m. The camera's pose is sampled at 12° intervals. The cameras have the same horizontal and vertical viewable angles, ϕ_h, ϕ_v, of 60° and have a working distance, r, of 7 m. The maximum visibility angle, θ_{max}, (2) and the angle, φ_{max}, (3) are taken to be 70° and 10° respectively. The pixel size of our vision system, p, is 0.02 mm and focal length, f, is 1.21 cm. The maximum stereo baseline length, $B_{max,}$, is 1.5 m. The cubic centre is located at the centroid of tetrahedron and each edge is 10 mm.

This case study illustrates how the variable stereo baseline length, camera positions and poses vary according to the accuracy requirement and the target location. In order to illustrate the cameras' positions and poses, the analysis considers the target model at four locations, 1, 2, 3 and 4, see Fig. 2. The arrows indicate the optical axes of cameras. The index numbers indicate the model locations and corresponding cameras' positions and poses calculated according to the maximum accuracy factor. The circles are the camera's potential sample positions. The sample positions and intervals are changed according to the camera's previous position with the constraint (9). In each position every upward triangular is visible to at least one stereo pair. The algorithm proves that a set of two pairs is sufficient to cover the three triangle surfaces. The stereo baseline length is dynamically changing according to the distance to the target.

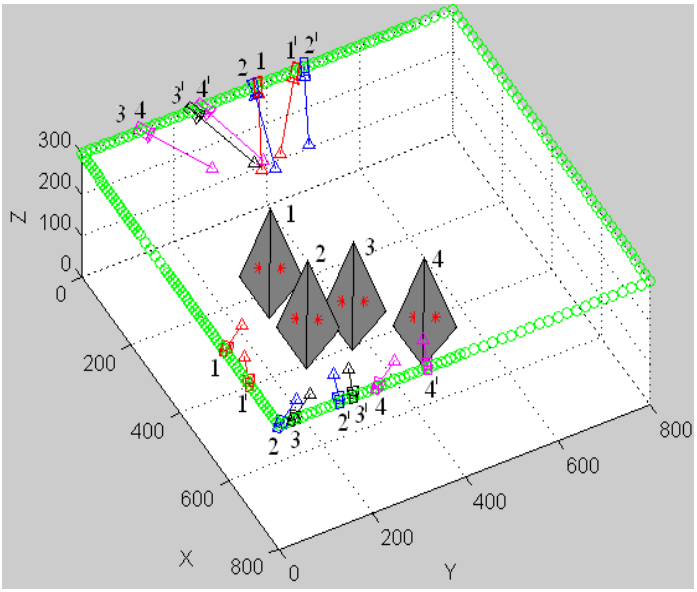


Fig. 2. The stereo pair positions, poses and baselines with the depth accuracy factor AF_{con} = 8, for the moving target.

Fig. 3 illustrates a case of four different values of AF_{con} applying to a target at the same position. The index number indicates the corresponding stereo pair according to AF_{con}. The stereo baseline lengths and reconstruction errors for the different accuracy factors are shown in Table I. It proves that the baseline increases as AF_{con} become more restricted, and the reconstruction error is smaller.

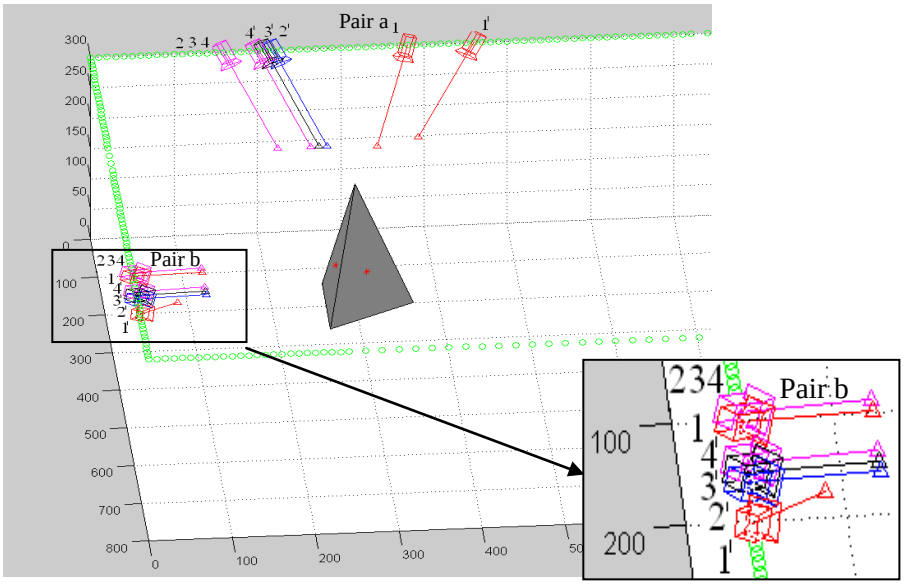


Fig. 3. The stereo pair positions and baseline lengths for the same target location vary according to the different accuracy.

Table I. The baseline lengths of two pairs and the reconstruction errors for different accuracy factors

IN	AF_{con}	B_a	B_b	Δ_{rec}	ΔZ_{max}
1	8	100	80	3.3	4.5
2	11	70	60	4.5	6.2
3	14	60	50	4.9	7.9
4	17	50	40	5.2	9.7

IN: the index number; AF_{con}: the reconstruction accuracy requirement; B_a, B_b: baseline lengths for each pair, [cm]; Δ_{rec}: the maximum value reconstruction error of pairs, [cm]; ΔZ_{max}: the theoretical maximum depth error, [cm].

5. Conclusion

The proposed approach is useful in the dynamic control of stereo pairs' baselines, and their corresponding positions and poses, to observe a human body and its activities space. The baseline of stereo pair may be adjusted according to the accuracy requirements and target distance. The planning and control of multi stereo pair's baselines, positions and poses for surveillance and tracking in supermarkets, museums and the home environment, and especially in situations which require stereo data to reconstruct 3D with a required accuracy, are possible fields of application.

The analysis of key factors which affect the accuracy of 3D reconstruction shows that the convergence angle, ψ, and target distance, Z, are most significant. The depth accuracy constraint may be sufficient to control the stereo pair's baseline length, position and pose. It is an effective method for system decision making and is easy to implement. From the simulation results, it is readily noticeable that the cubic reconstruction is useful in verifying the reconstruction accuracy and the proposed method of baseline length control has been proven. The two stages camera's position sampling has the flexibility to adjust the intervals and position ranges, and speed up computation. Future work could be to improve the human geometry model, for instance, it is possible to mark two of the upper triangles of tetrahedron to represent the human forefront. Camera zooming and human self-occlusions would be also an interesting study subject.

References

[1] E. Hörster and R. Lienhart, "On the Optimal Placement of Multiple Visual Sensors", *ACM International Workshop on Video Surveillance & Sensor Networks*, 2006.

[2] K. Chakrabarty, S.S. Iyengar, H. Qi and E. Cho, "Grid Coverage for Surveillance and Target Location in Distributed Sensor Networks", *IEEE Transaction on Computers*, 51(12): 1448-1453, 2002.

[3] J. Chen, S. Khatibi and W. Kulesza, "Planning of A Multi Stereo Visual Sensor System for A Human Activities Space", *the 2nd Int. Conf. on Computer Vision Theory and Applications*, 2007.

[4] A. Mittal, "Generalized Multi-Sensor Planning", *the 9th European Conference on Computer Vision*, 2006.

[5] E. Samson, D. Laurendeau, M. Parizeau, S. Comtois, J. Allan and C. Gosselin, "The Agile Stereo Pair for Active Vision," *Machine Vision and Application*, 17(1): 32-50, 2006.

[6] Y. Nakabo, T. Mukai, Y. Hattori, Y. Takeuchi and N. Ohnishi, "Variable Baseline Stereo Tracking Vision System Using High-Speed Linear Slider," *IEEE Int. Conf. on Robotics and Automation*, 2005.

[7] A. Fusiello, E, Trucco and A. Verri, "A Compact Algorithm for Rectification of Stereo Pairs," *Machine Vision and Applications,* 12: 16-22, 2000.

[8] P. N. M. Berkelaar and K. Eikland, "Lpsolve: Open Source (mixed-integer) Linear Programming System," *http://tech.groups.yahoo.com/group/lp_solve/files/,* Eindhoven University of Technology, 2005.

[9] G. L. Mariottini and D. Prattichizzo, "The Epipolar Geometry Toolbox: Multiple View Geometry and Visual Servoing for Matlab," *Proceedings of the IEEE Int. Conf. On Robotics and Automation,* 2005.

Paper VII

Evaluation of GM-PHD Filter for Multi-Target Tracking with a Stereo Vision System

Authors:

Jiandan Chen, Soheil Ghadami and Wlodek Kulesza

Reformatted version of paper originally accepted to:

IEEE International Instrumentation and Measurement Technology Conference, Hangzhou, China, May, 2011

Evaluation of the GM-PHD Filter for Multi-Target Tracking with a Stereo Vision System

Jiandan Chen, Soheil Ghadami and Wlodek Kulesza

Abstract

This paper evaluates the performance of the Gaussian Mixture Probability Hypothesis Density (GM-PHD) filter for tracking multiple targets in an intelligent vision system. A stereo vision camera is used to get the left and right image sequences in order to extract 3-D coordinates of the targets' positions in the real world scene. The 3-D trajectories of the targets are tracked by a GM-PHD filter. Moreover, the label continuity of the targets is guaranteed by a new method of labeling. Motion speed and angular velocity are proposed for the evaluation of the accuracy and label continuity of the filter in the implemented 3-D test motion model. The simulation results for two moving targets show that the proposed system not only robustly tracks them, but also maintains the label continuity of the two targets.

Keywords: *Performance Evaluation, Probability Hypothesis Density, Stereo Vision, Trajectory Tracking.*

1. Introduction

The tracking of moving objects has very diverse applications in almost all endeavours of life in present-day society, including security, surveillance, robotics, aeronautics, medicine, and sports. Object tracking is the fulcrum of the Intelligent Vision Agent System, IVAS, [1], which is a high-performance autonomous distributed vision and information processing system. The objective of multiple-target tracking is to estimate the number of targets at each time step from noisy measurements and correctly track them in the consecutive time steps. However, missed detection, vague tracks, and information lost due to occlusion and crossing-targets are among the essential problems in the motion tracking research field.

The *inferential measurement* is used to determine the dynamic behaviours of a process such as the ability of a system to process, store, transform, and transmit data, [2]. Validation and evaluation methods are the inferential measurement used to verify correctness and evaluate the performance of the analytical model and the procedure employed for a specific task. Evaluation results allow an estimation of the quality, reliability, efficiency, complexity and consistency of the analytical results. The measures used for evaluation are based on simulation and/or real experimental tests, [3].

In this paper, a parallel stereo vision system, a setup of two cameras with parallel optical axes, is applied to get the left and right image sequences of the targets' positions in the scene. The left camera's optical center is set as the original coordinate system, i.e. *xy*, plane is parallel to the left camera's image plane, and z is the targets' distance to the

left camera's position. Here, the proposed GM-PHD filter tracks the x, y, and z coordinates of the targets in the world coordinate system with respect to the left camera coordinate system, in contrast to many single-view filters which track the pixel coordinates in the image plane, [4]. The experimental results show that the proposed GM-PHD filter, combined with a stereo vision camera is able to properly track each dimension of the targets and associate observations with the appropriate targets. The Wasserstein distance from theoretical statistics was adopted as a means of defining a metric for multi-target distances which penalizes when the estimate of the number of targets is incorrect, [5]. The 3D circular motion test signal is introduced to evaluate the tracking performance according to the target speed and the angular velocity. The test signal with its standardised characteristics makes it possible to compare different methods and their efficiency for multi-target tracking

1.1 Survey of related works

Most traditional tracking algorithms such as Multiple Hypothesis Tracking (MHT), [6] and Joint Probabilistic Data Association Filter (JPDAF), [7], involve explicit association between measurements and targets. However, researchers have recently focused on Random Finite Sets (RFS) theory in which measurements and states are treated as random sets. Modeling set-valued states and set-valued observations as RFSs allows the problem of dynamically estimating multiple targets in the presence of clutter and association uncertainty to be cast in a Bayesian filtering framework, [8], [9]. A typical tracking algorithm based on the Bayesian's recursion computes the posterior probability density of a process based on the prior probability density of the process and the likelihood function, [8].

Furthermore, the multiple-target recursive Bayes filter based on the random Finite Set Statistics (FISST) is the theoretically optimal approach to multiple-sensor multiple-target detection, tracking, and identification, [10]. Mahler, [9], developed an approximation model called the Probability Hypothesis Density (PHD) filter for Bayes filtering to compensate for its computational intractability. A generic sequential Monte Carlo technique estimates states from the particles representing the posterior intensity using clustering techniques such as K-mean or expectation minimization, [11]. Cardinalized Probability Hypothesis Density (CPHD) is a generalization of the PHD recursion that jointly propagates the posterior intensity and the posterior distribution of the number of targets, [12].

Zhou and Aggarwal, [13], classified labeled tracked blobs into several model cues and the occluded target is matched with an appropriate class. Mozerov et al., [14], have proposed an approach based on the matching of the multiple trajectories in time to overcome long-term occlusion of targets. It is an interpolation algorithm to match the disconnected parts of the same trajectory during the occlusion. Parrilla et al., [15], presented an algorithm for 3-D tracking objects in a stereo video sequence by combining optical flow and stereo vision, and then proposed an adaptive filter and neural networks to handle occlusion. Darrell et al., [16], combined stereo, colour, and face detection modules into a single system and tracked face and body parts.

1.2 Problem statement and main contribution

A combination of a tracking filter with a stereo camera extends the vision system ability to track the multi target in 3-D space. Target movement features are influential factors in the evaluation of the tracking filter performance. The target speed motion and angular velocity play key roles in the GM-PHD filter's tracking performance.

The main contributions of the paper can be summarized as follows:

- The GM-PHD filter for multi-target tracking using a stereo vision camera is modeled and implemented in MATLAB, and validated by applying two random motion trajectories;
- The 3-D spatial motion test signal is introduced to evaluate the filter tracking performance in terms of its accuracy and label continuity.
- The two key parameters, the motion speed $r\omega$, and angular velocity ω, in the motion model test signal are proposed for the evaluation of the GM-PHD filter's tracking performance;
- The robustness of the filter performance on occlusions for multitracking is investigated.

2. Modeling of the GM-PHD filter and the test spatial motion introduction

2.1 A model of the GM-PHD filter for multi-target tracking

In the framework of the multi-target tracker, the multi-target state can be described by the RFS. The multi-target state can be represented as a discrete time k set $\mathbf{X}_k$ defined as:

$$\mathbf{X}_k = \left\{ \chi_{k,i} : i = 1, \cdots, M_\chi(k) \right\} \quad , \tag{1}$$

where $M_\chi(k)$ is the number of targets in the scene in time k, and i is the index variable.

The multi-target measurement $\mathbf{Z}_k$ from the camera sensor is the set:

$$\mathbf{Z}_k = \left\{ \zeta_{k,j} : j = 1, \cdots M_\zeta(k) \right\} \quad , \tag{2}$$

where $M_\zeta(k)$ is the number of observations in the time k, and j is the index variable.

The GM-PHD filter is based on three additional assumptions compared to the PHD filter [8]:

(i) Each target follows a linear Gaussian dynamical model, and the sensor has a linear Gaussian measurement model that can be modeled as:

$$\begin{aligned} f_{k|k-1}(\chi|\varsigma) &= N(\chi; \mathbf{F}_{k-1}\varsigma; \mathbf{Q}_{k-1}) \quad , \\ g_k(\zeta|\chi) &= N(\zeta; \mathbf{H}_k\chi; \mathbf{R}_k) \quad , \end{aligned} \tag{3}$$

where $N(\cdot; m; \Psi)$ denotes a Gaussian density with the mean and the co-variance Ψ, $\mathbf{F}_{k\text{-}1}$ is the state transition matrix, $\mathbf{Q}_{k\text{-}1}$ is the process noise co-variance, $\mathbf{H}_k$ is the observation matrix, and $\mathbf{R}_k$ is the observation noise co-variance.

(ii) The survival and detection probabilities denoted as $p_{S,k}(\chi) = p_{S,k}$, $p_{d,k}(\chi) = p_{d,k}$ are independent.

(iii) The intensity of the spontaneous birth is a Gaussian mixture can be expressed as:

$$\gamma_k(\chi) = \sum_{i=1}^{J_{\gamma,k}} w_{\gamma,k}^{(i)} N\left(\chi; m_{\gamma,k}^{(i)}, \Psi_{\gamma,k}^{(i)}\right) \quad , \tag{4}$$

where $J_{\gamma,k}$, $w_{\gamma,k}^{(i)}$, $m_{\gamma,k}^{(i)}$, $\Psi_{\gamma,k}^{(i)}$, i=1,…, $J_{\gamma,k}$ are given model parameters that determine the shape of the birth intensity [4]. The posterior intensity at time k-1 is given as follows:

$$v_{k-1}(\chi) = \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} N(\chi; m_{k-1}^{(i)}, \Psi_{k-1}^{(i)}) \quad . \tag{5}$$

Then, the predicted intensity at time k is given by:

$$v_{k|k-1}(\chi) = v_{s,k|k-1}(\chi) + \gamma_k(\chi) \quad , \tag{6}$$

where γ is the birth intensity, w is the weight parameter, and:

$$v_{S,k|k-1}(\chi) = p_{S,k} \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} N(\chi; m_{S,k|k-1}^{(i)}, \Psi_{S,k|k-1}^{(i)}) \quad . \tag{7}$$

After the object detection has been finished and $\mathbf{Z}_k$ is available, the posterior intensity is updated by:

$$v_k(\chi) = (1 - p_{d,k}) v_{k|k-1}(\chi) + \sum_{\zeta \in \mathbf{Z}_k} \sum_{i=1}^{J_{k|k-1}} w_k^{(i)}(\zeta) N(\chi; m_{k|k}^{(i)}, \Psi_{k|k}^{(i)}) \quad . \tag{8}$$

2.2 Track labeling in the GM-PHD filter

With the stereo vision camera's configurations given in Chapter IV and taking the left camera's coordinate system of the stereo pair as the reference one, we are able to obtain the reconstructed position $\zeta_{k,j} = (x_{k,j}, y_{k,j}, z_{k,j})$ of the target j at the time k, in the real world frame with respect to the left camera's coordinate system, [17], where:

$$\begin{cases} z_{k,j} = \dfrac{bf}{D_{k,j}\Delta} \\ x_{k,j} = \dfrac{u_{l_{k,j}} z_{k,j}}{f} \\ y_{k,j} = \dfrac{v_{l_{k,j}} z_{k,j}}{f} \end{cases} , \tag{9}$$

and k and j subscripts are skipped for generalization; (u_l, v_l) and (u_r, v_r) are the left and right image plane coordinates in pixels, respectively, Δ is the pixel size, $D = u_l - u_r$ is the disparity value, and f and b are the focal length and baseline length respectively.

Each target is assigned by a fix label. If a new target appears, a new label is added to the set, and similarly if a target dies or disappears, its corresponding label is discarded from the labels' set. The set of labels at time k is $L_k = \{l_k^1, l_k^2, l_k^{Jk}\}$, where J_k is the number of targets at time k.

Consider $\mathbf{Z}_k = \{\zeta_{k\text{-}1,\ 1}, \zeta_{k\text{-}1,\ 2}\}$, a two-target measurement set, and its corresponding prediction state $\hat{\mathbf{Z}}_{k|k-1} = \{\hat{\zeta}_{k|k-1,1}, \hat{\zeta}_{k|k-1,2}\}$. The labels' set is defined as a two-value set, i.e, $L_k = \{l_k^1, l_k^2\}$, in which l_k^1 and l_k^2 are labels of the first and the second target, respectively. The filter detects the targets' label discontinuity if the following condition holds:

$$\left\| \zeta_{k-1,1} - \hat{\zeta}_{k|k-1,1} \right\| > \left\| \zeta_{k-1,2} - \hat{\zeta}_{k|k-1,1} \right\| \quad . \tag{10}$$

Thus, the targets' labels at time k are swapped to maintain label continuity of the targets.

2.3 Modeling of the test spatial motion

The point $\zeta_{k-1,j} = (x_{k-1,j}, y_{k-1,j}, z_{k-1,j})$ at the time k-1 represents the target j position in the 3-D space as shown in Fig. 1. This point is assumed to have moved to a new position $\zeta_{k,j}$ at the time k. The target motion can be described as follows:

$$\zeta_{k,j} = \zeta_{k-1,j} + \Delta k V_{k-1,j} \alpha_{k-1,j} \quad , \tag{11}$$

where $\alpha_{k-1,j} = (\cos\theta_{k-1,j}, \sin\theta_{k-1,j}, \sin\varphi_{k-1,j})$ is the direction vector of velocity of target j in the 3-D space. The angle $\theta_{k-1,j} \in [0, 2\pi]$, is the angle between motion direction and YZ plane, and $\varphi_{k-1,j} \in [0, 2\pi]$ is the angle between the motion direction and the XY plane. $V_{k-1,j}\,\alpha_{k-1,j}$ denotes the velocity in the time k-1, and $V_{k-1,j}$ is the speed (absolute amplitude of velocity). The sample interval, Δk, is the observation sampling time when the target moves from position $\zeta_{k-1,j}$ to $\zeta_{k,j}$.

In order to evaluate the performance of the proposed GM-PHD filter, a 3D circular test motion is proposed and also shown in Fig. 1. The target's initial position ζ_0 in the 3-D space is described by:

$$\zeta_{0,j} = O + r_j\left(\cos\alpha_{0,j}, \sin\alpha_{0,j}, \sin\beta_{0,j}\right) \quad , \tag{12}$$

where $O=(o_x, o_y, o_z)$ is the center of the test motion, r_j is the radius from O to the target's initial position, α_0 is an angle between the radius vector and YZ plane, $\alpha_0 \in [0, 2\pi]$, and β_0 is an angle between the radius vector and XY plane, $\beta_0 \in [0, \pi]$, .Then, the target motion at position $\zeta_{k,j}$ can be approximated by:

$$\zeta_{k,j} = \zeta_{k-1,j} + r\omega\Delta k\left(\cos\theta_{k-1,j}, \sin\theta_{k-1,j}, \sin\varphi_{k-1,j}\right) \quad , \tag{13}$$

where the angular velocity is $\omega = 2\pi/(K\cdot\Delta k)$, and K is the measurement sampling rate in samples/period. The motion direction alteration is defined as $\Delta\theta_{k,j} = \omega\Delta k$. The motion speed is $r\omega$. Fig. 2 shows the image plane view and top view of two targets following the test motion model.

3. Validation and performance evaluation of the GM-PHD filter

As mentioned previously, the two images are taken by a pair of CCD cameras modeled as pinhole cameras. The simulated space considers a regular room of the size 8×8×3 m, in which targets are covered by the cameras' Field of View. The Epipolar Geometry Toolbox, [18], is used to project the target's position in the 3-D space to the image planes of the size 1024×1024 pixels. Each camera of the stereo vision system has a focal length of 6 mm and a pixel size of 12.9 μm. The baseline length is 100 mm. A Gaussian noise distribution with the variance of half a pixel size is used during measurements.

A random walk motion is used to validate the implementation of the GM-PHD filter. In addition, the 3-D spatial test motion is used to evaluate the performance of the filter. The mean value of absolute error defined as the difference between the prediction and the ground truth was measured by using the Wasserstein distance, [19].

3.1 Validation of GM-PHD filter for multi-target tracking

A random walk motion evaluates the filter's multi-target tracking ability. The random trajectories are generated according to (11) with 70 samples. The target motion speed is

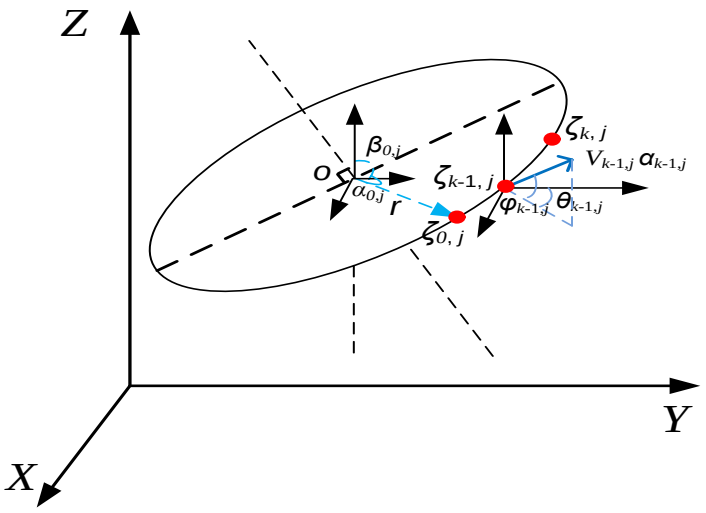


Figure 1. Target circular motion illustration

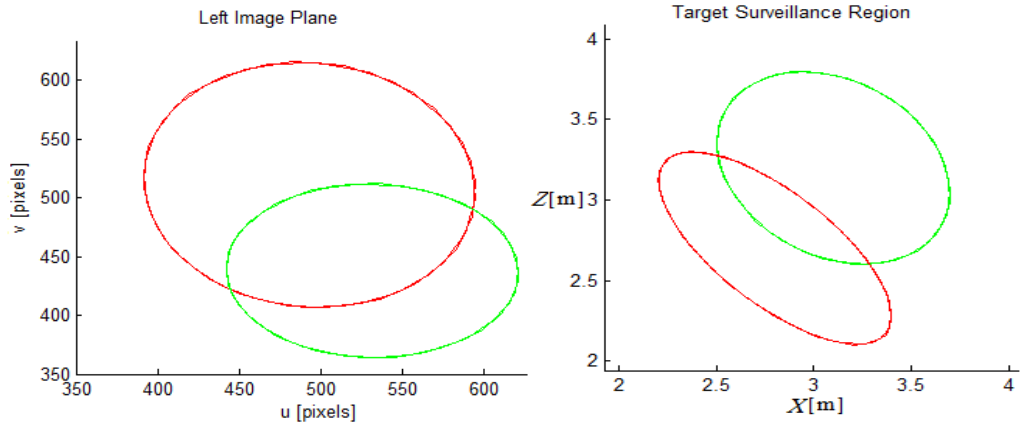
generated by Gaussian distribution in MATLAB. Two targets are observed within the surveillance region.

Fig. 3 indicates the ability of the stereo vision based GM-PHD filter to track targets in each coordinate. The mean values of the absolute error in *x*, *y*, and *z* dimensions are 66.3 mm, 19.6 mm, and 110.6 mm, respectively. The two targets' labels are well preserved through the tracking process.

3.2 Performance evaluation of the GM-PHD filter

The spatial test motion defined by (11), (12) and (13) is applied to evaluate the performance of the tracker when two targets move in the space, causing occlusions and crossings to occur.

In order to evaluate the potential of the GM-PHD filter, two targets whose trajectories follow the spatial motion model with the equal value r=1.4 m, and K=25 samples/period in two periods are applied. Fig. 4 shows that the filter tracks two targets quite well, and



(a) (b)

Figure 2. (a) Left image plane view and (b) top view of the scene. Red paths with motion parameters $\alpha_{0,1}$ = 180 °, $\beta_{0,1}$ =135 °, r_1=0.6 m, and green paths with motion parameters $\alpha_{0,2}$ = 180 °, $\beta_{0,2}$ =135 °, r_2=0.6 m.

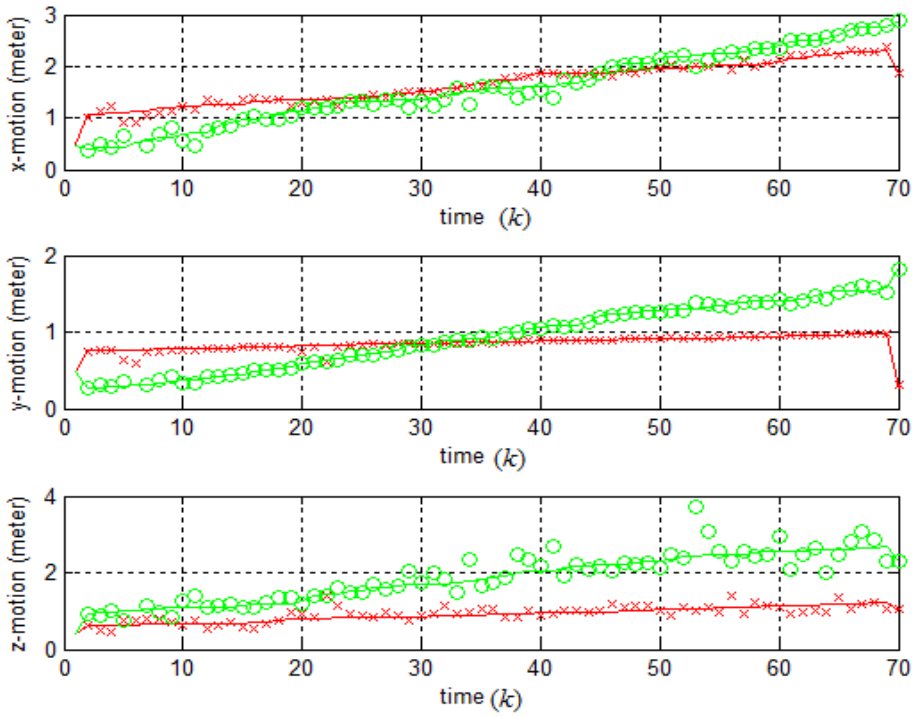


Figure 3. The target motion path in the world coordinate system, and the ground truth is marked as the solid lines, and the predictions produced by the filter are marked as crosses and circles.

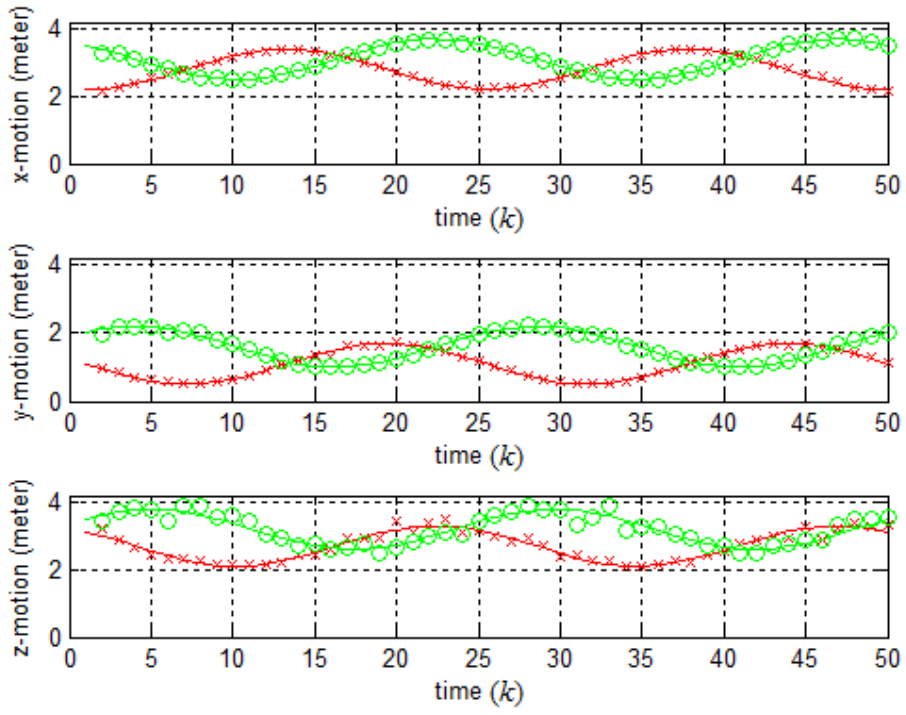


Figure 4. The filter maintains labels and while tracking targets in two periods. The ground truth is marked as the solid lines, and the predictions by the filter are marked as crosses and circles.

that the targets' identities are maintained even when two targets cross and occlude each other.

The ability of the filter to track targets with increasing speed motion is evaluated by changing the radius *r*. Table I shows the mean and variance values in the *x*, *y*, and *z* dimensions for different radiuses and *K*= 25 samples/period in fifty periods. The mean value of the absolute error is proportional to the radius *r*.

Fig. 5 and Fig. 6 show that the mean value of the absolute error in the *x* and *y* dimensions decreases when the sampling rate *K* increases for different values of radius, respectively. The black, green, and red curves correspond to radius $r = 1.4$ m $r = 1$ m, $r = 0.6$ m, respectively. *K* varies from 5 samples/period to 40 samples/period.

It is shown in Fig. 7 that the mean value of absolute error in the *z* dimension decreases with the increase of the sampling rate. However, the mean error is higher than the other two dimensions. It may come from an inaccurate depth reconstruction due to spatial noise and the estimation of camera's intrinsic and extrinsic parameters. The most influential noise in each camera is the spatial noise, which may originate from discrete sensor pixels and the resolution of the focal length. Furthermore, the depth reconstruction accuracy depends on the system configuration which is defined by sensor resolution, focal length, and baseline length. However, depth spatial quantization plays a key role in determining the accuracy of a 3-D reconstruction, [1]. The effect of reconstruction on the tracking performance is not discussed in this paper.

Table I. Mean and variance of the tracking error in each dimension for K=25 samples/period in fifty periods

Radius *r* [m]	***x***		***y***		***z***	
	Mean [mm]	Var [mm]	Mean [mm]	Var [mm]	Mean [mm]	Var [mm]
0.60	32.40	0.049	33.10	0.18	121.60	1.21
1.00	41.20	0.031	48.90	0.12	143.20	0.56
1.40	56.70	0.020	66.20	0.31	159.30	1.41

4. Conclusion

A new approach to tracking multiple targets by the GM-PHD filter in the real scene with the use of a stereo vision system is presented. It makes possible to handle occlusion by gathering 3-D information of the activity space. It is shown that the filter successfully tracks targets while maintaining each target's label and performs well even when one target is occluded by the other one, or when they cross each other.

A 3-D test motion of the spatial movement is modeled by two key parameters, motion speed and angular velocity, which are useful for the evaluation of the tracking filter. The angular velocity and speed of the test motion model are easily related to the sampling rate and the radius what simplifies the filter evaluation.

The error between the prediction and the ground truth in the tracking shows that the GM-PHD filter performance depends on target motion features and the tracking accuracy is proportional to motion speed and angular velocity.

In future research, it would be useful to evaluate the performance of the GM-PHD filter using multiple cameras in multiple-view multiple-target tracking problems. The compassion of performance for different tracking algorithm can be implemented.

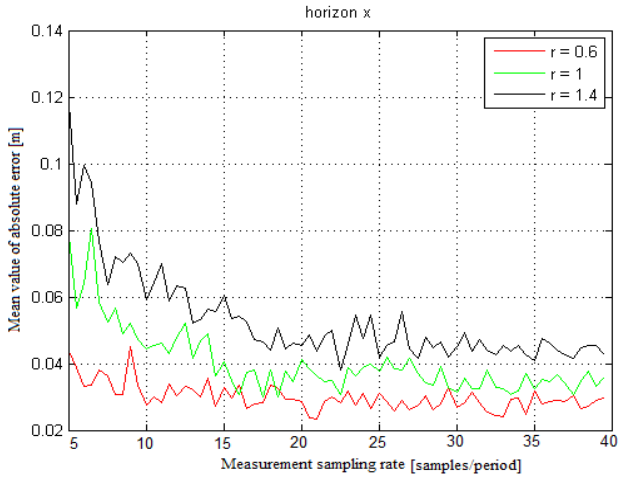


Figure 5. The mean value of the absolute error for different radius vs. sampling rate (samples/period) in x dimension.

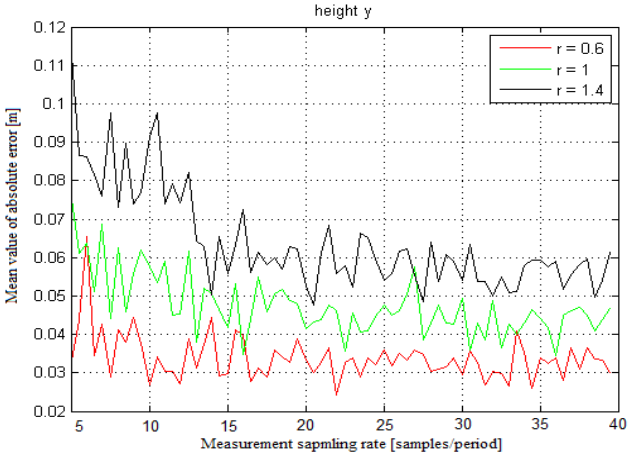


Figure 6. The mean value of the absolute error for different radius vs. the sampling rate (samples/period) in y dimension.

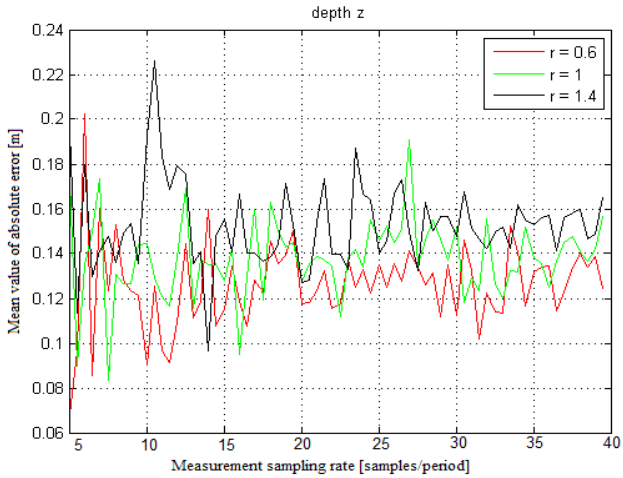


Figure 7. Mean value of the absolute error for different radius vs. the sampling rate (samples/period) in z dimension.

References

[1] J. Chen, S. Khatibi, and W. Kulesza, “Planning of a multi stereo visual sensor system- depth accuracy and variable baseline approach”, in: Proc. of IEEE Computer Society 3DTV-Con, the True Vision Capture, Transmission and Display of 3D Video, Kos, Greece, 2007.

[2] J. McGhee, V. Henderson, M. J. Korczynski, W. Kulesza, Scientific Metrology, Lodart S.A., Lodz, Poland, 1st ed, Reprint with corrections, 1996.

[3] J. Wirandi, Modelling and Validation of Industrial Measurement Systems-Aspects of Quality and Human Factors, PhD thesis, Lund University, 2007.

[4] J. Chen, O. E. Adebomi, O. S. Olusayo, and W. Kulesza, “The evaluation of the Gaussian mixture probability hypothesis density approach for multi-target tracking”, IEEE Int. Conference on Imaging Systems and Techniques, Greece, 2010.

[5] J. Hoffman and R. Mahler, “Multi-target miss distance via optimal assignment”, IEEE Trans. Sys., Man, and Cybernetics-Part A, vol. 34, no. 3, pp. 327-336, 2004.

[6] S. Blackman, “Multiple hypothesis tracking for multiple target tracking”, IEEE Aerosp. Electron. Syst. Mag., vol. 19, no. 1, pp. 5-18, 2004.

[7] Y. Bar-Shalom and T. E. Fortmann, Tracking and Data Association, San Diego, CA, USA, 1988.

[8] B. N. Vo and W. K. Ma,”The Gaussian mixture probability hypothesis density filter”, IEEE Transaction on Signal Processing, vol. 54, no. 11, pp. 4091-4104, 2006.

[9] R. Mahler, “Multi-target Bayes filtering via first order multi-target moments”, IEEE Trans. Aerosp. Electron. Syst., vol. 39, no. 4, pp. 1152–1178, 2003.

[10] Z. Hongjian, J. Zhongliang, and H. Shiqiang, “Tracks extraction of the probability hypothesis density filter for survival targets”, in: Proceedings of the 27th Chinese Control Conference, Kunming, China, pp. 343-347, 2008.

[11] B. Vo, S. Singh, and A. Doucet, “Sequential monte carlo implementation of the PHD filter for multi-target tracking” in: Proc. of Int. Conf. Inf. Fusion, Cairns, Australia, pp. 792–799, 2003.

[12] R. Mahler, “PHD filters of higher order in target number”, IEEE Transactions on Aerospace and Electronic Systems, vol. 43, no. 4, pp. 1523 – 1543, October 2007.

[13] Q. Zhou and J. K. Aggarwal,”Object tracking in an outdoor environment using fusion of features and cameras”, Image and Vision Computing, vol. 24, no. 11, pp. 1244-1255, 2006.

[14] M. Mozerov, A. Amato, and X. Roca, “Trajectory occlusion handling with multiple-view distance minimization clustering”, Optical Engineering, vol. 47, no. 4, 2008.

[15] E. Parrilla, J. Reira, J. Torregrosa, and J. Hueso, “Handling occlusion in object tracking in stereoscopic video sequences”, Elsevier, Journal of Mathematical and Computer Modelling, vol. 50, no. 5-6, pp. 823-830, Sep 2009.

[16] T. Darrell, G. Gordon, M. Harville, and J. Woodfill, “Integrated person tracking using stereo, color, and pattern detection”, International Journal of Computer Vision, vol. 37, no. 2, pp. 175-85, 2000.

[17] A. Dankers, N. Barnes, E. Zelinsky, “Active vision, rectification and depth mapping”, Australian Conference on Robotics and Automation, 2004.

[18] G. L. Mariottini and D. Prattichizzo, “The epipolar geometry toolbox: multiple view geometry and visual servoing for MATLAB”, In: Proc. Of the IEEE Int. Conf. on Robotics and Automation, 2005.

[19] D. Clark, K. Panta and B.-N. Vo, “The GM-PHD filter multiple target tracker”, 9th International Conference on Information Fusion, July 2006.

Paper VIII

An Adaptive Quality Assessment System – Aspect of Human Factor and Measurement Uncertainty

Authors:

Jenny Wirandi, Jiandan Chen and Wlodek Kulesza

Reformatted version of paper originally published in:

IEEE Transactions on Instrumentation and Measurement, vol. 58, no. 11, pp. 68-75, January, 2009

An Adaptive Quality Assessment System – Aspect of Human Factor and Measurement Uncertainty

Jenny Wirandi, Jiandan Chen and Wlodek Kulesza

Abstract

In this article, we discuss a model of quality that makes use of the fuzzily defined variable approach in order to better understand the concept and thus enable the further development of this variable. We propose a general method that may estimate a quality index which handles both qualitative and quantitative issues. The system furthermore uses a neural network, since the system learns how to integrate the human factor into a quantitative quality index. In our case study, we have examined the measurement of image quality and proposed a theoretical model of pulp quality.

Keywords: *Fuzzily Defined Quantity, Neural Network, Image Quality, Pulp Quality, Quality Index, Quality Assessment.*

1. Introduction

Often, the evaluation of scientific and industrial measurements fails due to a lack of traceable calibration of the specific type of measurements or instruments being used. For example, when the measured variable is fuzzy – an *attribute*, such as *product quality* or *smell* or a *service* – the traceability chain related to calibration may be incomplete or missing altogether, [1], [2]; the systematic error cannot be reduced, and the *Guide to the Expression of Uncertainty in Measurement*, (GUM), [3], cannot be applied, [4].

An additional issue is the modelling problem of the *Fuzzily Defined Variable*, (FDV), since the FDV often consists of both the quantitative and the qualitative factors, which are of different importance for different targets or users. Ordinarily, the way to solve the problem of describing the FDV *quality* is to create an index that depends only on quantitative features. Some qualities have a defined and accepted common index, but some quality systems have to be continuously updated due to varying requirements, particularly those qualities composed of both quantitative and qualitative factors.

Moreover, it has been indicated that the lack of a clearly defined quality index, e.g., pulp quality, reduces employee awareness of the product quality, [5].

In this article, we propose a general *adaptive* system that defines and estimates a quality index which can merge both qualitative and quantitative features and factors. The system can be updated continuously and adapted for time and for changeable

features, [6]. The *Neural Network*, (NN), is a tool suited to the system, since it can learn how to integrate both qualitative factors and quantitative features.

In our case studies, we apply two different fuzzily defined variables, *image quality* and *pulp quality*. The main difference between those variables is that image quality often has a reference – a pattern image – that can be compared with and this is not the case with pulp.

The article is structured as follows: in the next section, we present research that relates to the subject. In section 3, we present the formulation of the problem and the main uses of the fuzzily defined variable. Section 4 presents the results of our two case studies and the implementations of image quality and pulp quality. Finally, in section 5 we discuss our findings and end with some concluding remarks.

2. Related works

In order to estimate and compare quality, different kinds of *Quality Indices*, (QI), have been developed in different branches such as the food industry, the pulp industry, ecology and image processing.

In the mid 1980s, the *Tasmanian Food Research Unit in Australia* developed a method called *Quality Index Method* (QIM), [7]. Later on, this method was further improved by European fisheries research institutes. The QIM is based on well-defined characteristic changes occurring in the odour and texture of eyes, skin and gills of raw fish. An advantage of the QIM is that the quality index increases linearly with storage time in ice. Moreover, the information may be used in production management, [8].

The food industry has implemented a *Quantitative Descriptive Analysis*, (QDA), which defines the sensory attributes of food. These sensory attributes include texture, odour and flavour and the QDA thus provides a detailed description of all qualitative and quantitative attributes. A trained panel is given a broad selection of reference samples, and, under the guidance of a panel chairman, they use the samples to create a terminology that describes all detectable aspects of the product, [9].

In the ecological field, a *Water Quality Index*, (WQI), is used to define criteria for the classification of surface water. The index is based on characterisation data for large quantities of water, [10], [11]. It applies a weight factor according to the *relevance* of the feature as an indicator of water quality, [11].

In the paper pulp industries, there are company specific quality indices, but so far no defined index has been adapted by the whole industry. Pulp quality needs to be defined from case to case, since it has no standardised definition and since quality demands change according to what it should be used for, [4], [5], [12]. In fact, there is indication that employees within the same company define pulp quality in different ways, [5]. At the same time, papermakers combine different measurement features to determine the quality of pulp. Their choice of features depends on what kind of paper they produce. Their decision is furthermore influenced by the tradition/education of the employees and of the company. The pulp manufacturer has to select these features in order to satisfy all customers and enable them to make their own assessment of the product, [5].

To return to image quality, there are several multidimensional aspects to consider when assessing this. There are different image quality indices, depending on the application area. In general, there are two dissimilar views of how to define image quality. These views relate to whether an existing pattern image is available or if physical limits are clearly defined, [13]. A review of several ways of describing image quality was carried out by Engeldrum, who states that the most successful ones are the Minkowski metric and other metrics related to Minkowski, [13]. A new image quality index is proposed by Wang *et al.*, [14], [15]. It relies on three quantitative features: structured distortion, luminance distortion and contrast distortion. Their quality index is defined mathematically, and the input measurement is based on the difference between a reference image and the measured image. It has been indicated that the index correlates with the human visual system and thus with human assessment.

3. Problem formulation and a proposed solution

The FDV attributes are not clearly defined since they depend on different types of features and factors. The choice of these features and factors depends on the customer/target group, and/or the cultural environment, and/or the tradition of education within the application field. Since this FDV often depends on both quantitative and qualitative factors, it can not be expressed in only quantitative terms, [4]. Due to this, the two main dependencies that must be handled within the FDV are related to:

1. The set of features that are a part of the FDV, and depend on:
   a. *Expertises;*
   b. *Possible measurements;*
   c. *Pattern data.*
2. The weights on the FDV and depend on:
   a. *Human perception - assessment;*
   b. *The feature's relevance;*
   c. *Measurement uncertainty;*
   d. *Other factors such as cost or complexity.*

The quality index may thus be calculated with the help of the different methods described in the introduction. The index should aid the quality assessment in a comparable and more objective way. We propose an index that is created through a method that uses the NN, which could be generalised for many different kinds of products and services. The method uses (for more details see Fig. 3 and Fig. 9 in section IV):

1. A set of quantitative features which can be re-selected;
2. A set of quantitative factors which can be re-selected;
3. A set of qualitative factors, which are used to train the system.

Fig. 1 illustrates the modelling of quality. The *initial quality model* is established by experts in the field. The set of *quantitative features* to be included in the QI and their initial weights, [α], are based on the *measurement uncertainty* and *relevance* of each

feature. Then, the *adaptive quality model* is trained to integrate the relationship between the value of the *quantitative features* and the *human subjective assessments* of different types of products or services.

When the product or service is used for different purposes, the human assessment can differ completely. In these cases, a group classification method based on *Principal Components Analysis*, (PCA), is useful. The judges are grouped according to different factors that may determine how they subjectively assess the product's quality. Such factors include the purposes of the product, job positions, background, etc. The applied group classification procedure is as follows:

- In order to remove the non-significant components, the PCA is applied before evaluating the QI.
- The *Root Mean Square Deviation*, (RMSD), values of the reconstructed quantified assessments are calculated for all the possible groups.
- The groups are recognised as being distinguishable if the RMSD value is greater than the discretisation step of the neural network index. Otherwise, the groups cannot be distinguished.

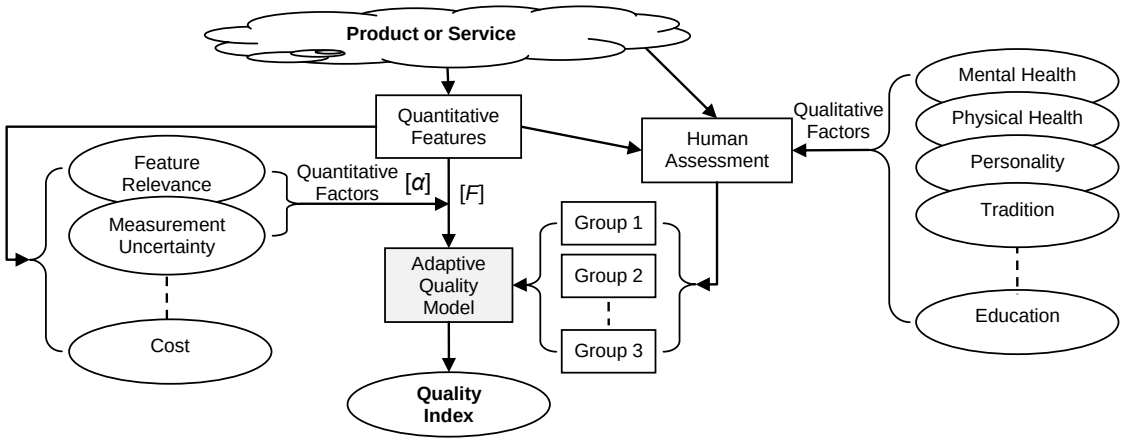


Fig. 1. A block diagram illustrating the quality model. Ellipses denote representations of information, and rectangles denote process transformation from one representation into another.

In the training stage, two input data, quantitative human assessment and the quantitative features, train the NN. This stage requires several epochs of training to adjust the NN-weights to meet the output performance goal, [16]. Then, the trained quality model estimates the discrete QI of the product/service based on both the quantitative features and factors and the knowledgeable human assessments.

The modelling procedure can be summarised by recounting the following steps:

1. Definition of the *initial quality model* with a selection of input quantitative features, [F], and quantitative factors, represented by weights [α].

2. *Group classification*, by finding the correlation between human assessment and qualitative factors.
3. *Training stage* for self-organising NN input layers according to classified groups and estimation of the weights of NN.
4. *Validation stage*, to get the *discrete* QI.

4. Case studies and results

The quality model described in section 3 has been implemented for *image* and *pulp*. The QIs have been estimated by a simplified image quality system. The validation stage which occurs after the training stage of the adaptive system using NN, is shown in Fig. 2. The system classifies the qualitative factors into different target groups. Then, the NN estimates the QI for each target group based on the quantitative features and factors.

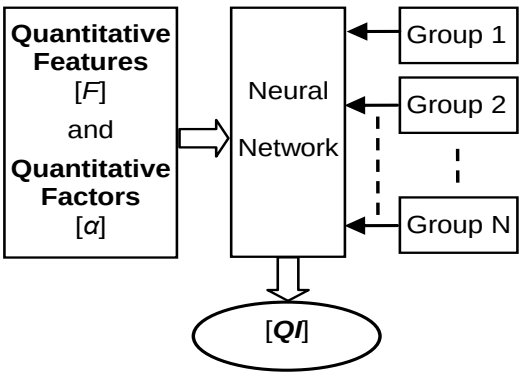


Fig. 2. The validation stage of the adaptive quality system.

The weights are calculated from quantitative factors. For instance, they could be calculated as a mean value or product of the measurement uncertainty ratio, α_{ui} and relevance ratio, α_{ri}, e.g.:

$$\alpha_i = \frac{\alpha_{ui} + \alpha_{ri}}{2}, \tag{1}$$

where the measurement uncertainty ratio α_{ui} is calculated as:

$$\alpha_{ui} = 1 - u_c(y_i), \tag{2}$$

where $u_c(y_i)$ is the combined relative standard uncertainty of output estimates, y_i , [3].

The relevance ratio, α_{ri}, could be calculated as the ratio of a numbers of experts who recognise the feature, i, as relevant, h_i, to a total number of experts, H.

$$\alpha_{ri} = \frac{h_i}{H}. \tag{3}$$

4.1 Application to image quality

The image quality is influenced by quantitative features such as *basic properties*, *naturalness* and *colourfulness*. The initial weight of each quantity is estimated from the quantitative factors' *measurement uncertainty*, *cost*, *data* and *relevance*. However, the human quality assessment depends also on many qualitative factors such as *personal background*, *physical environment*, *usefulness*, *tools* and *pattern representation*, which are related to the target and the human being. Therefore, image quality can be modelled by the Ishikawa diagram presented in Fig. 3.

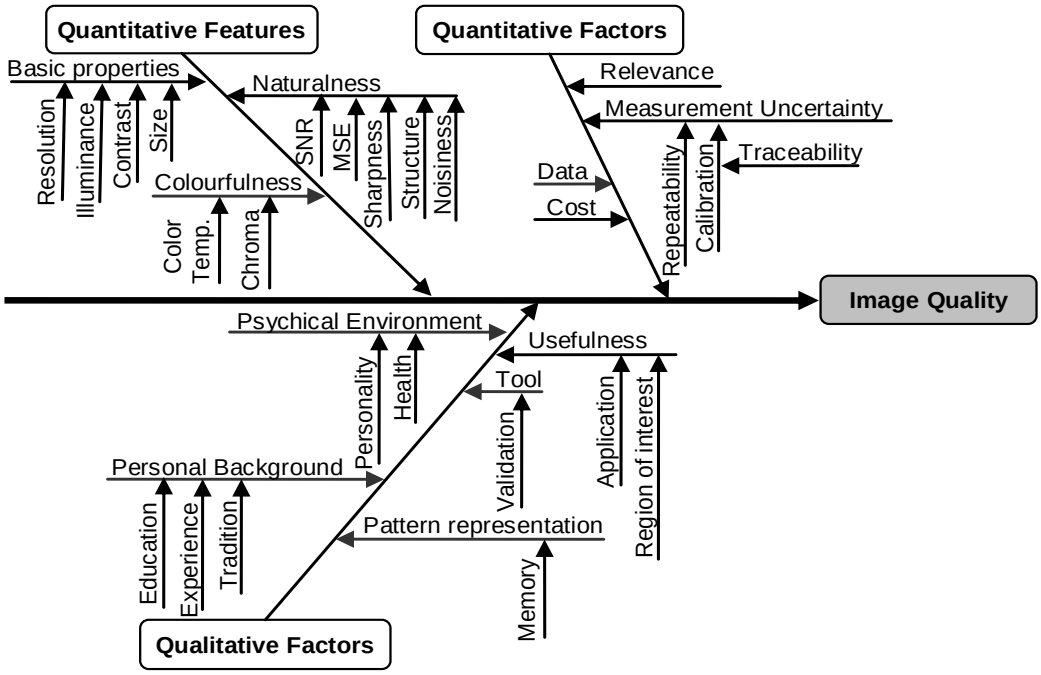


Fig. 3. Definition of the image quality modeled by the Ishikawa diagram.

4.1.1 Initial quality model

The simplified quality model has been tested by estimating the QI of greyscale images. The NN consists of two stages: a training stage and an validation stage. In the training stage, three types of quantitative features are used in the quality model: *structure distortion ratio*, F_s, and two basic properties, *luminance distortion ratio*, F_l, and *contrast distortion ratio*, F_c. These measurements of the intensity data of the pattern and test images are normalised, [14], [15]. This can be described in the following way:

$$F_s = \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \tag{4}$$

$$F_l = \frac{2\bar{x} \cdot \bar{y}}{\bar{x}^2 + \bar{y}^2}, \tag{5}$$

$$F_c = \frac{2\sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2}, \tag{6}$$

where x denotes the *intensity data* of the image being tested, and y denotes the *intensity data* of the image pattern; σ_x, σ_y are the standard deviations related to the *contrast* of the test and pattern images respectively; σ_{xy} is the covariance representing *structure similarity*.

Since all features are normalised, they are less than or equal to one, $|F_i| \leq 1$. Moreover, the product of these three weighted features, F_{IQ}, is implemented as a fourth feature, where the respective weights, α:s, are based on the measurements' uncertainty and features' relevance, [15]:

$$F_{IQ} = (F_s)^{\alpha_s} \cdot (F_l)^{\alpha_l} \cdot (F_c)^{\alpha_c}. \tag{7}$$

For reasons of simplification, in this case study all weights, α:s, are considered to equal one unit.

4.1.2 Group classification and training stage

To formulate the relationship between the quantitative features and the qualitative factors, 51 people were asked to assess 21 training images with the same illustration, but with different types and levels of distortions, such as *gaussian noise*, *salt-pepper noise*, *multiplicative speckle noise*, *blurring* and *JPEG compression distortion*, (see example in Fig. 4). They were asked to assess the quality of the images on a scale of 0 to 9, where 0 represents the best quality. They were informed that the images were not to be ranked and that several images could have the same grade. Histograms of the collected data from our questioning show that they are normally distributed.

In the first step of the procedure, we chose to classify the images with different groups of people. The groups's assessments could be biased due to gender and/or because they may have had previous experience with image processing. The reconstructed quantified assessments were computed from the first three principle components resulting from applying PCA. Next, the mean values of the reconstructed grades were taken for each image within each group.

The male and female groups consisted of 27 and 24 people respectively. The mean reconstruction quantified assessments for each image for both groups are presented in Fig. 5. The RMSD of those two groups' quantified assessments for all 21 images is 0.36. Fig. 6 shows the mean reconstruction quantified assessments for groups according to their experience with image processing. The groups of experienced and non-experienced people consisted of 20 and 31 persons respectively. The RMSD of these two groups' quantified assessments of 21 images is 0.46. Both RMSD values, based on gender and on experience, are less than 1, which is the discretisation step of NN QI. In this case, we can conclude that the different groups provide compatible quality assessments. Therefore, distinguishing people participating in this study in different groups is unnecessary.

The NN was implemented with the help of the *Matlab Neural Network Toolbox* and the three-layer *transig*/*transig*/*logsig* network with ten neurons in each layer, [16]. Afterwards, the *Back-Propagation approach Neural Network* was applied. The training function *trainlm* was used, and the training procedure had to be repeated up to 5000 epochs to meet the output performance goal with a *Mean Square Error* of 0.02. The training stage was then considered complete.

4.1.3 Validation stage

After the training stage, the NN was used to estimate image QIs for a *new* set of 21 images with a different illustration, (see examples in Fig 7). This validation of the image quality index based on the images' quantitative features should anticipate the groups' assessments. To validate the model, the new images were also evaluated by 15 people. The results given by the model approximates the human assessment within the QI resolution (see Fig. 8). Furthermore, based on this result, one can conclude that the model recognizes different kinds of disturbance.

When we compare the results of the QIs generated by our adaptive quality system with the results calculated using the method proposed by Wang *et al*., [14], [15], it is clear that our QIs predict the human assessment with much better accuracy, (see Fig. 8).

4.2 Application to pulp quality

Another promising application field of the quality model is pulp quality assessment. Pulp has many different application areas. Therefore, to be able to properly derive a relationship between QI and human quality assessment, it is necessary to first classify the involved people into tentative groups depending on the application at hand. To obtain the relationship between the quantitative features and the qualitative factors shown in Fig. 9, producers and/or customers have to assess training samples. Because of the different target groups, a classification method based on PCA, which groups people according to the subjective quantitative assessment and qualitative factors, is required, (see Fig. 2).

4.2.1 Initial quality model

The set of essential quantitative quality features that are a part of the *Chemical Pulp Quality* model includes measurements of *single fibre properties*, *physical pulp properties*, *cleanliness* and *colour*, [4], [5], [12]. The pulp quality can be defined by the Ishikawa diagram presented in Fig. 9.

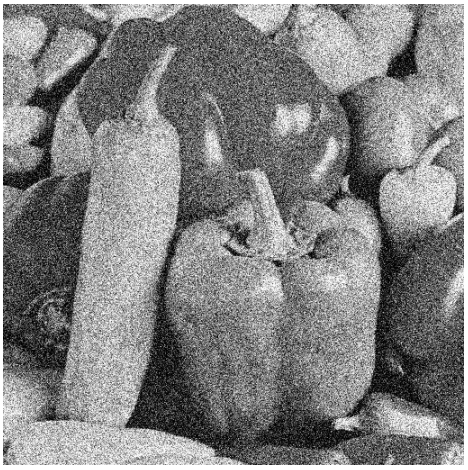


Fig.4a. The training image distorted by Gaussian noise and human grade=6.54.

Fig.4b. The training image distorted by blurring distortion and human grade=5.47.

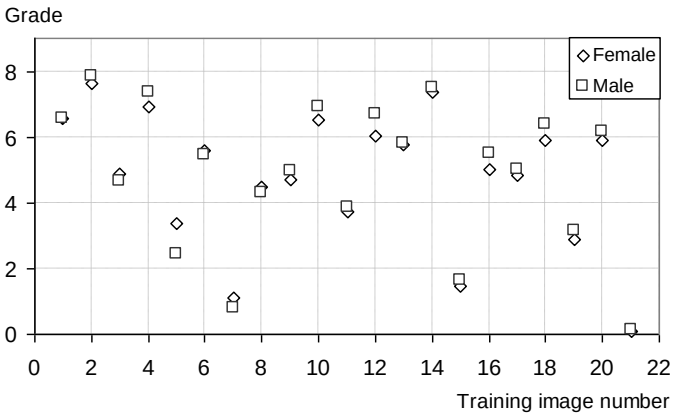


Fig. 5. The mean reconstruction quantified assessments for each image for the male and female groups.

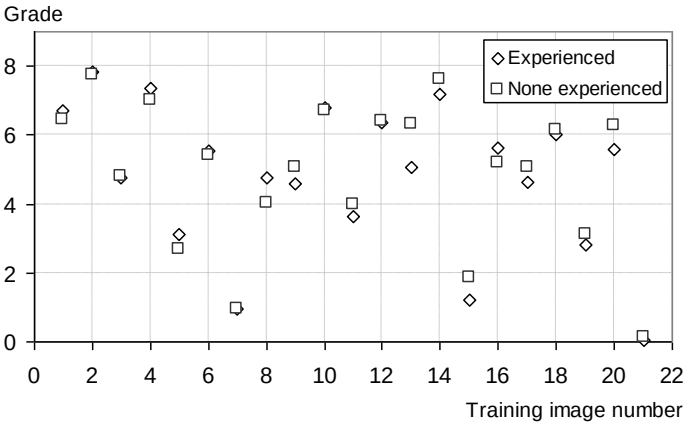


Fig. 6. The mean reconstruction quantified assessments for each image for the two groups, experienced and none experienced with image processing.

Fig. 7a. Test image distorted by Gaussian noise: QI=7 and human grade=7.14. This image is represented by the number 3 in Fig. 8.

Fig. 7b. Test image distorted by salt-pepper noise: QI=4 and human grade=3.57. This image is represented by the number 15 in Fig. 8.

Fig. 7c. Test image distorted by multiplicative speckle noise: QI=8 and human grade=6.07. This image is represented by the image number 7 in Fig. 8.

Fig. 7d. Test image distorted by JPEG compression distortion: QI=5 and human grade=5.57. This image is represented by the image number 9 in Fig. 8.

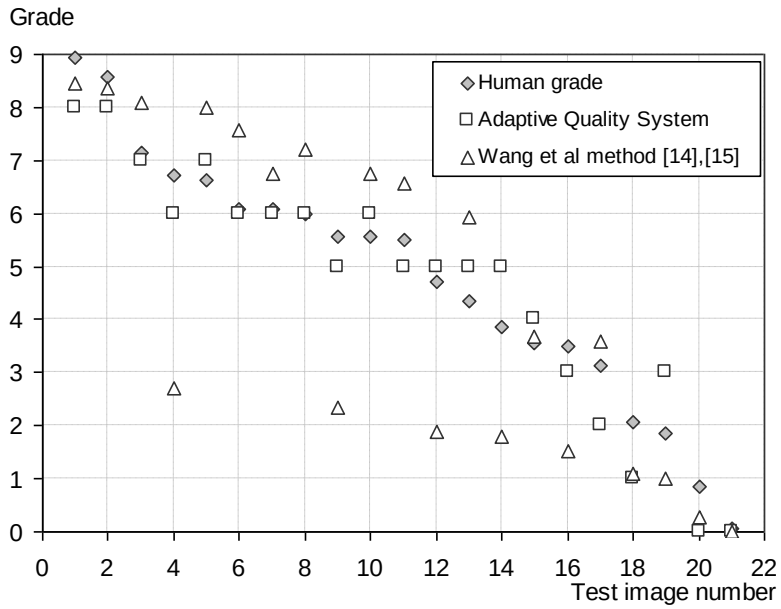


Fig. 8. A validation of the image quality system. The grade of the images ranges from 0 to 9, where 0 represents the best quality.

In our case study, we implemented three essential features into the quality model: fibre length, *f*, tear index, *t*, and brightness, *b*. The features were chosen based on a poll regarding pulp quality conducted at the company where the model was tested, [5]. These features are mathematically modelled as quantitative features by implementing the method presented in [14], [15]. Since a reference pattern does not exist for the pulp, the pulp companies usually apply features such as target value (for our case study, respectively l_{target}, t_{target} and b_{target}), as goals for the respective measurements of fibre length, tear index and brightness. Apart from the target values, upper and lower permissible target values for fibre length and tear index are also used. However, when measuring brightness, only the target value and the lower permissible target value need be taken into account, since measurement results above the target value are harmless.

Based on the limits discussed above, the *standard target deviations* for the respective features could be determined. It is important to note that since different pulp products are produced within the same factory, the preferred target value and upper/lower limits vary according to the manufactured product.

The target value and standard target deviation of each feature, together with measurements of samples, form the normalised features. Each of the three features can be expressed as two quality elements, one related to the mean values of the relevant number of samples, $F_l^{(1)}, F_t^{(1)}, F_b^{(1)}$ respectively, and one related to standard deviation, $F_l^{(2)}, F_t^{(2)}, F_b^{(2)}$ respectively:

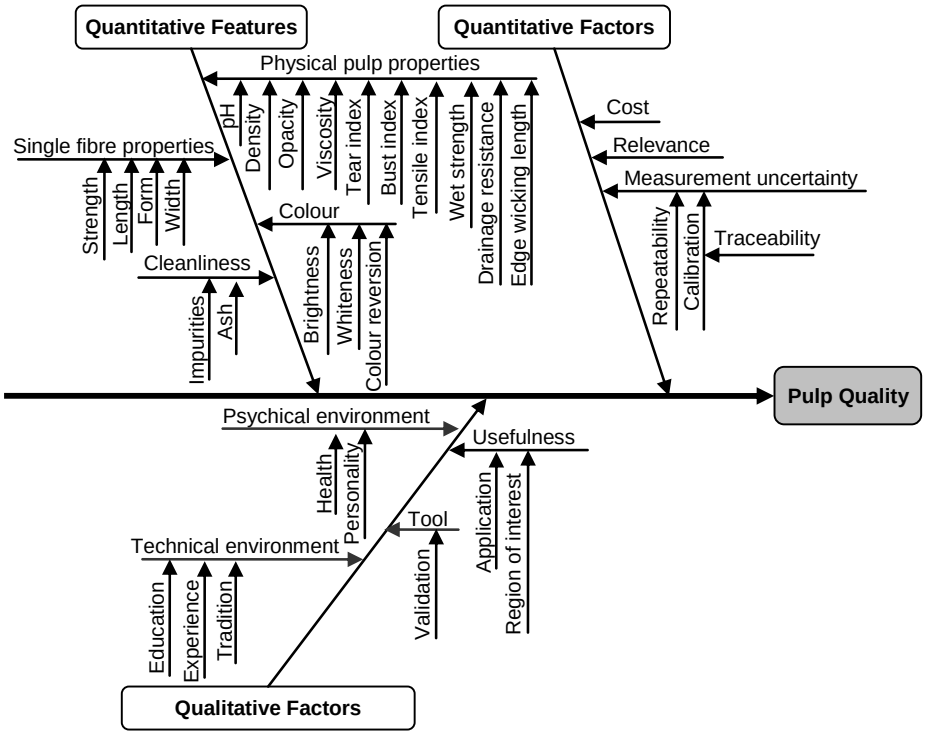


Fig. 9. Definition of the pulp quality modelled by the Ishikawa diagram.

$$F_l^{(1)} = \frac{2\bar{l} \cdot l_{target}}{\bar{l}^2 + l_{target}^2}, \tag{8}$$

$$F_l^{(2)} = \frac{2\sigma_l \cdot \sigma_{l_{target}}}{\sigma_l^2 + \sigma_{l_{target}}^2}, \tag{9}$$

$$F_t^{(1)} = \frac{2\bar{t} \cdot t_{target}}{\bar{t}^2 + t_{target}^2}, \tag{10}$$

$$F_t^{(2)} = \frac{2\sigma_t \cdot \sigma_{t_{target}}}{\sigma_t^2 + \sigma_{t_{target}}^2}, \tag{11}$$

$$F_b^{(1)} = \frac{2\bar{b} \cdot b_{target}}{\bar{b}^2 + b_{target}^2}, \tag{12}$$

$$F_b^{(2)} = \frac{2\sigma_b \cdot \sigma_{b_{target}}}{\sigma_b^2 + \sigma_{b_{target}}^2}, \tag{13}$$

where σ_l, σ_t, σ_b are the standard deviations for test fibre length, tear index and brightness respectively; $\sigma_{l_{target}}$, $\sigma_{t_{target}}$, $\sigma_{b_{target}}$ are the standard target deviations for target fibre length, tear index and brightness respectively; $\bar{l}$, $\bar{t}$ and $\bar{b}$, are the mean values of fibre length, tear index and brightness respectively, and l_{target}, t_{target}, b_{target} are the target values of fibre length, tear index and brightness respectively.

Since all features are normalised, they are less than or equal to one, $\left|F_i^{(j)}\right| \le 1$. As with the image quality model, the product of all exponentially weighted features, F_{PQ}, is used:

$$F_{PQ} = \left(F_l^{(1)}\right)^{\alpha_l^{(1)}} \cdot \left(F_l^{(2)}\right)^{\alpha_l^{(2)}} \cdot \left(F_t^{(1)}\right)^{\alpha_t^{(1)}} \cdot \left(F_t^{(2)}\right)^{\alpha_t^{(2)}} \cdot \left(F_b^{(1)}\right)^{\alpha_b^{(1)}} \cdot \left(F_b^{(2)}\right)^{\alpha_b^{(2)}}, \quad (14)$$

where $\left|\alpha_i^{(j)}\right| \le 1$.

The weights, $\alpha_i^{(j)}$:s, of each feature in (14) depend on the quantitative factors *measurement uncertainty*, *cost* and *relevance*. The weights are also evaluated by experts. Since the model results strongly depend on the chosen weights, $\alpha_i^{(j)}$, they need most likely to be adjusted several times during the modelling phase before the most appropriate values can be achieved. This is the case, since there is indication that the employees within a pulp company attribute varying degrees of *relevance* to different quantitative pulp quality features, [5].

5. Conclusions and discussion

The proposed QI model can handle both qualitative and quantitative factors, as well as the features that are a part of the FDV. The model focuses on the human assessment of product or service quality. As a modelling tool, the NN is used.

The proposed objective group classification method is useful in cases when the assessment of different customer/user groups differs significantly.

The calculation of the weight of each quantitative feature should be based on its measurement uncertainty and on its relevance for the quality assessment. Please note that the measurement feature's relevance should be based on hard facts as much as possible as this minimizes the errors associated with human assessment. Furthermore, it is important that each factor be considered only once. For example, a lack of traceable calibration lies within the bounds of measurement uncertainty and should not be a part of the analysis of the feature's relevance. The Ishikawa diagram is a valuable tool since it can help avoid such errors during the modelling process.

The human assessment of the product's quality is a part of the model training stage during which the weights and the layers of NN are established.

The model has been tested and validated with the help of the image QI. The QI estimated by the adaptive system and human quantitative assessments matched each other very well. Our model of image quality was tested on the same image, but with different disturbances. The results could be improved by testing the system with a more significant number of images, as well as with different kinds of illustrations and a bigger group of people.

We have also presented a QI model of pulp where some quality features related to fibre length, tear index and brightness have been defined and used in the adaptive quality model of pulp.

A common definition of pulp quality will encourage pulp quality awareness, which in turn will increase a given company's competitive strength. However, to be able to make use of the quality index's full potential within the company, it is necessary to educate and regularly update the employees on how to influence the measurement features essential to product quality.

Finally, to be able to adopt the QI completely, it is necessary to be aware of the fact that people interpret measurement data differently. It is important to educate employees regarding how they, in their daily work, may influence the measurement features essential to product quality and to regularly update their knowledge of these features. The continual education of staff dealing with measurements at different stages of development and production is crucial for improving product quality. The methods of education must be flexible and adapted to different levels and requirements.

We believe that the idea introduced in this paper is a promising starting point for the future development of a representative QI that allows for a more efficient comparison of products or services based on both quantitative and qualitative factors. If developed further, research along the proposed lines may well lead to potential dynamic calibrations of the Fuzzily Defined Variable – *Quality*.

Acknowledgements

The authors wish to acknowledge Dr. Siamak Khatibi for his valuable discussions, and Paul Curley and Johan Höglund for their comments.

The main thesis of this paper was presented at the *2007 IEEE International Workshop on Advanced Methods for Uncertainty Estimation in Measurement* in Sardagna, Trento, Italy, July 16-18, 2007.

References

[1] J. Wirandi, A. Lauber, "Reliability of fibre measurements", in *Proc. World Pulp and Paper Week, SPCI International Conference*, Stockholm, Sweden, June, 2005, pp. 47-52.

[2] J. Wirandi, A. Lauber, W. Kulesza, "Problem of applying modern uncertainty concepts to the measurement of instrument specific parameters", *IEEE Trans. Instrum. Meas.*, vol. 55, no. 3, pp. 700-705, 2006.

[3] ISO. *Guide to the Expression of Uncertainty in Measurement*. International Organization for Standardization. Geneva, Switzerland, 1995.

[4] J. Wirandi, A. Lauber, "Uncertainty and traceable calibration – how modern measurement concepts improve product quality in process industry", *Elsevier Meas.*, vol. 39, no. 7, pp. 612-620, 2006.

[5] J. Wirandi, W. Kulesza, A. Lauber, "Human factor validation in an industrial measurement system", *Elsevier Meas.*, vol.41, no. 7, pp. 705-718, 2008.

[6] J. Wirandi, J. Chen, W. Kulesza, "An Adaptive Model of the Fuzzy Variable - Quality Index", in: *Proc.* Int. Workshop AMUEM, Trento, Italy, July 16-18, 2007.

[7] A. Bremner, "A convenient easy to use system for estimating the quality of chilled seafood", *Fish Processing Bulletin*, no. 7, pp. 59-73, 1985.

[8] E. Martinsdottir, K. Sveinsdottir, J. Luten, R. Schelvis-Smit, G. Hyldig, *Sensory Evaluation of Fish Freshness*, QIM Eurofish, 2001.

[9] H. Stone, J.L. Sidel, "Quantitative descriptive analysis: developments, applications, and the future", *Food Technology*, vol. 52, no. 8, pp. 48-52, 1998.

[10] Couillard, Y. Lefebvre, "Analysis of water quality indices", *Journal of Environment Management*, vol. 21, no. 2, pp. 161-179, 1985.

[11] N. Stambuck-Giljanovic, "Water quality evaluation by index in Dalmatia", *Water Research*, vol. 33, no.16, pp. 3423-3440, 1999.

[12] J.-E. Levlin, L. Söderhjelm, *Pulp and Paper Testing*. Gummerus Printing, book 17, Helsinki, Finland, 1999.

[13] P.G. Engeldrum, "Image quality modeling: where are we?" in: *Proc. IS & T's PICS Conference*, pp. 251-255, 1999.

[14] Z. Wang, A.C. Bovik, "A universal image quality index", *IEEE Trans. Signal Processing Letters*, vol. 9, no. 3, pp. 81-84, 2002.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.

[16] The MathWorks, "Neural Network Toolbox 5.0.1." (May, 2007). [Online]. Available at: http://www.mathworks.com/products/neuralnet.

ABSTRACT

In our society with its aging population, the design and implementation of a high-performance distributed multi-sensor and information system for autonomous physical services become more and more important. In line with this, this thesis proposes an Intelligent Multi-Sensor System, IMSS, that surveys a human activities space to detect and identify a target for a specific service. The subject of this thesis covers three main aspects related to the set-up of an IMSS: an improved depth measurement and reconstruction method and its related uncertainty, a surveillance and tracking algorithm and finally a way to validate and evaluate the proposed methods and algorithms.

The thesis discusses how a model of the depth spatial quantisation uncertainty can be implemented to optimize the configuration of a sensor system to capture information of the target objects and their environment with required specifications. The thesis introduces the dithering algorithm which significantly reduces the depth reconstruction uncertainty. Furthermore, the dithering algorithm is implemented on a sensor-shifted stereo camera, thus simplifying depth reconstruction without compromising the common stereo field of view.

To track multiple targets continuously, the Gaussian Mixture Probability Hypothesis Density, GM-PHD, algorithm is implemented with the help of vision and Radio Frequency Identification, RFID, technologies. The performance of the tracking algorithm in a vision system is evaluated by a circular motion test signal. The thesis introduces constraints to the target space, the stereo pair characteristics and the depth reconstruction accuracy to optimize the vision system and to control the performance of surveillance and 3D reconstruction through integer linear programming. The human being within the activity space is modelled as a tetrahedron, and a field of view in spherical coordinates are used in the control algorithms.

In order to integrate human behaviour and perception into a technical system, the proposed adaptive measurement method makes use of the Fuzzily Defined Variable, FDV. The FDV approach enables an estimation of the quality index based on qualitative and quantitative factors for image quality evaluation using a neural network.

The thesis consists of two parts, where Part I gives an overview of the applied theory and research methods used, and Part II comprises the eight papers included in the thesis.

Keywords: 3D Reconstruction, Depth Measurement, Depth Reconstruction, Dither, Human Factor, Image Quality, Iso-disparity Surfaces, Multi-Sensor System, Quality Measurement, Sensor Arrangement, Surveillance, Tracking, Uncertainty.

ISSN 1653-2090
2011:05
ISBN: 978-91-7295-200-3