ABSTRACT

In our aging society, the design and implementation of a high-performance autonomous distributed vision information system for autonomous physical services become ever more important. In line with this development, the proposed Intelligent Vision Agent System, IVAS, is able to automatically detect and identify a target for a specific task by surveying a human activities space. The main subject of this thesis is the optimal configuration of a sensor system meant to capture the target objects and their environment within certain required specifications. The thesis thus discusses how a discrete sensor causes a depth spatial quantisation uncertainty, which significantly contributes to the 3D depth reconstruction accuracy. For a sensor stereo pair, the quantisation uncertainty is represented by the intervals between the iso-disparity surfaces. A mathematical geometry model is then proposed to analyse the iso-disparity surfaces and optimise the sensors' configurations according to the required constraints. The thesis also introduces the dithering algorithm which significantly reduces the depth reconstruction uncertainty. This algorithm assures high depth reconstruction accuracy from a few images captured by low-resolution sensors.

To ensure the visibility needed for surveillance, tracking, and 3D reconstruction, the thesis introduces constraints of the target space, the stereo pair characteristics, and the depth reconstruction accuracy. The target space, the space in which human activity takes place, is modelled as a tetrahedron, and a field of view in spherical coordinates is proposed. The minimum number of stereo pairs necessary to cover the entire target space and the arrangement of the stereo pairs' movement is optimised through integer linear programming.

In order to better understand human behaviour and perception, the proposed adaptive measurement method makes use of a fuzzily defined variable, FDV. The FDV approach enables an estimation of a quality index based on qualitative and quantitative factors. The suggested method uses a neural network as a tool that contains a learning function that allows the integration of the human factor into a quantitative quality index.

The thesis consists of two parts, where Part I gives a brief overview of the applied theory and research methods used, and Part II contains the five papers included in the thesis.

Keywords: 3D Reconstruction, Iso-disparity Surfaces, Depth Reconstruction Uncertainty, Uncertainty Analysis, Dither, Sensor Placement, Multi Stereo View, Image Quality, Human Factor.
A Multi Sensor System
for
a Human Activities Space
Aspects of Planning and Quality Measurement

Jiandan Chen
A Multi Sensor System
for
a Human Activities Space
Aspects of Planning and Quality Measurement

Jiandan Chen

Department of Signal Processing
School of Engineering
Blekinge Institute of Technology
SWEDEN
Abstract

In our aging society, the design and implementation of a high-performance autonomous distributed vision information system for autonomous physical services become ever more important. In line with this development, the proposed Intelligent Vision Agent System, IVAS, is able to automatically detect and identify a target for a specific task by surveying a human activities space. The main subject of this thesis is the optimal configuration of a sensor system meant to capture the target objects and their environment within certain required specifications. The thesis thus discusses how a discrete sensor causes a depth spatial quantisation uncertainty, which significantly contributes to the 3D depth reconstruction accuracy. For a sensor stereo pair, the quantisation uncertainty is represented by the intervals between the iso-disparity surfaces. A mathematical geometry model is then proposed to analyse the iso-disparity surfaces and optimise the sensors’ configurations according to the required constrains. The thesis also introduces the dithering algorithm which significantly reduces the depth reconstruction uncertainty. This algorithm assures high depth reconstruction accuracy from a few images captured by low-resolution sensors.

To ensure the visibility needed for surveillance, tracking, and 3D reconstruction, the thesis introduces constraints of the target space, the stereo pair characteristics, and the depth reconstruction accuracy. The target space, the space in which human activity takes place, is modelled as a tetrahedron, and a field of view in spherical coordinates is proposed. The minimum number of stereo pairs necessary to cover the entire target space and the arrangement of the stereo pairs’ movement is optimised through integer linear programming.

In order to better understand human behaviour and perception, the proposed adaptive measurement method makes use of a fuzzily defined variable, FDV. The FDV approach enables an estimation of a quality index based on qualitative and quantitative factors. The suggested method uses a neural network as a tool that contains a learning function that allows the integration of the human factor into a quantitative quality index.

The thesis consists of two parts, where Part I gives a brief overview of the applied theory and research methods used, and Part II contains the five papers included in the thesis.

**Keywords:** 3D Reconstruction, Iso-disparity Surfaces, Depth Reconstruction Uncertainty, Uncertainty Analysis, Dither, Sensor Placement, Multi Stereo View, Image Quality, Human Factor.
Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisors in no particular order: To my examiner Prof. Ingvar Claesson for giving me the opportunity to conduct the research I love at the Blekinge Institute of Technology, BTH, and for his supervision; to Prof. Włodek Kulesza for being a great mentor and for giving me important advice and crucial guidance during my research work. I sincerely appreciate the countless hours he has devoted to our papers and to my thesis. His efforts and contributions made this thesis possible. Furthermore, I am most thankful to Dr. Siamak Khatibi for his profound knowledge and experience in the field of computer vision, and I wish to extend my gratitude to him for engaging in many fruitful discussions and providing numerous creative ideas. Last, but definitely not least, I want to thank Dr. Benny Lövström who has supported and helped me through the whole research project. He also helps me enjoy life at BTH.

I would also like to thank my present colleagues at BTH and former colleagues at the University of Kalmar. The faculties of these departments made me feel at home. There is always such a pleasant atmosphere at the Department of Signal Processing.

I am most grateful to my wife, Haiyan, for her patience, endless love and support for my work. She moved with me to Sweden, stands by me always, and gives a lot of pleasure to our life. Very special thanks to my daughter Zihan who has given me the inspiration and the energy I needed to keep working. I also want to thank the rest of my family for their love and support.

I am very grateful to Johan Höglund, Akademiska Språkbyrån, Kalmar, Sweden and Paul Curley for their comments.

Ronneby, June 2008

Jiandan Chen
# Contents

ABSTRACT .................................................................................................................................... V  
ACKNOWLEDGEMENTS ........................................................................................................ VII  
ACRONYMS.................................................................................................................................. XI  
APPENDED PAPERS................................................................................................................ XIII  
1 INTRODUCTION..................................................................................................................... 1  
  1.1 Thesis objective ................................................................................................................ 3  
  1.2 Thesis scope ..................................................................................................................... 3  
  1.3 Research methods .......................................................................................................... 4  
  1.4 Thesis outline .................................................................................................................. 4  
2 DEPTH RECONSTRUCTION METHOD AND ACCURACY ............................................... 7  
  2.1 The iso-disparity surfaces geometry model .................................................................... 8  
  2.2 Depth reconstruction ..................................................................................................... 9  
3 ARRANGEMENT OF MULTI SENSORS FOR A HUMAN ACTIVITIES SPACE ........... 11  
  3.1 Constraints for the optimisation model ........................................................................ 11  
  3.2 Implementation of the camera planning with the integer linear programme ............ 14  
4 AN ADAPTIVE MEASUREMENT METHOD .................................................................... 15  
  4.1 An adaptive measurement method and its implementation ........................................ 15  
  4.2 The adaptive method for image quality measurement ................................................ 17  
  4.3 Conclusions and further development ......................................................................... 18  
5 SUMMARY ............................................................................................................................ 19  
  5.1 Summary of contributions ............................................................................................ 19  
  5.2 Conclusions .................................................................................................................. 21  
  5.3 Future research ............................................................................................................. 22  
REFERENCES ........................................................................................................................... 25  
PAPER I .......................................................................................................................................... 31  
PAPER II ........................................................................................................................................ 45  
PAPER III ....................................................................................................................................... 67
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC</td>
<td>Analogue to Digital Converter</td>
</tr>
<tr>
<td>AF</td>
<td>Accuracy Factor</td>
</tr>
<tr>
<td>AGP</td>
<td>Art Galley Problem</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge Coupled Device</td>
</tr>
<tr>
<td>FDV</td>
<td>Fuzzily Defined Variable</td>
</tr>
<tr>
<td>FoV</td>
<td>Field of View</td>
</tr>
<tr>
<td>GUM</td>
<td>Guide to the Expression of Uncertainty in Measurement</td>
</tr>
<tr>
<td>ILP</td>
<td>Integer Linear Programming</td>
</tr>
<tr>
<td>IR</td>
<td>Image Resolution</td>
</tr>
<tr>
<td>IVAS</td>
<td>Intelligent Vision Agent System</td>
</tr>
<tr>
<td>JPEG</td>
<td>Joint Photographic Experts Group</td>
</tr>
<tr>
<td>MIT</td>
<td>Massachusetts Institute of Technology</td>
</tr>
<tr>
<td>NN</td>
<td>Neural Network</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Components Analysis</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Distribution Function</td>
</tr>
<tr>
<td>PPM</td>
<td>Perspective Projection Matrix</td>
</tr>
<tr>
<td>QDA</td>
<td>Quantitative Descriptive Analysis</td>
</tr>
<tr>
<td>QI</td>
<td>Quality Indices</td>
</tr>
<tr>
<td>QIM</td>
<td>Quality Index Method</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root Mean Square Deviation</td>
</tr>
<tr>
<td>WQI</td>
<td>Water Quality Index</td>
</tr>
</tbody>
</table>
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:

**Paper I**


**Paper II**


**Paper III**


**Paper IV**


**Paper V**

Part I
1 Introduction

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 vision information system which can understand human behaviour and living environments, and thus temporarily substitute a qualified nurse and housekeeper.

Human-centred computation is proposed by the MIT Oxygen Project, [1]. The computation here is centred on human needs and abilities instead of on the needs and possibilities of the machine. Furthermore, Hashimoto has suggested the concept of intelligent space: Intelligent space 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, it is an interface both for the human and for robots.

The proposed Intelligent Vision Agent System, IVAS, is a high-performance autonomous distributed vision and information processing system. Figure 1.1 illustrates the idea of the IVAS. It consists of multiple sensors and actuators for surveillance of the human activities space which includes the human being and their surrounding environment, such as robots, household appliances, lights, and so on. The system not only gathers information, but also controls these sensors including their deployment and autonomous servo. The most important function, however, is to extract the required information from images for different applications, such as three dimension (3D) reconstruction, etc. The 3D information from a real scene of target objects can be compared with a pattern in order to make decisions. Meanwhile, the pattern may also be renewed by the inclusion of a learning phase. These features require the system to dynamically adjust the camera to get 3D information in an optimal way. 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.

Similar to the human eyes, stereo vision observes the world from two different points of view. At least two images need to be fused to obtain a depth perception of the world. However, due to the digital camera principle, the depth reconstruction accuracy is limited by the sensor pixel resolution and causes quantisation of the reconstructed 3D space. The spatial quantisation is illustrated by iso-disparity maps. The iso-disparity surfaces approach when calculating the reconstruction uncertainty has been discussed by Völpel and Theimer, [3]. In addition to this, the shape of iso-disparity surfaces for general stereo configurations was studied by Pollefeys et al., [4]. Furthermore, the
The proposed mathematical model of the iso-disparity map provides an efficient way to describe the shape of the iso-disparity planes and estimate a depth reconstruction uncertainty which is related to the stereo pair baseline length, the target distance to the baseline, the focal length, the convergence angle, and the pixel resolution.

The depth spatial quantisation uncertainty caused by a discrete sensor is one of the factors which influence the depth reconstruction accuracy the most. This type of uncertainty cannot be decreased by reducing the pixel size because of the restricted sensitivity of the sensor itself and because of the declining signal to noise ratio. The selection of an optimal sensor pixel is discussed by Chen et al., [5].

Dithering is a well-known technique that is applied in analogue to digital converters, ADCs. This method decreases the system resolution below the least significant bit, [6], [7], [8]. As shown in Paper II, the introduced spatial dithering signal can cause the depth spatial quantisation uncertainty to be reduced by half by combining four pairs of stereo images. The target space, specified as a human activities space, determines the planning of stereo sensors in order to increase the sensor observability for the optimal number of sensors.

The planning of the sensor by the use of reannealing software was introduced by Mittal, [9], and the evaluation of the sensors’ configurations by a quality metric was presented in [10]. Furthermore, a linear programming method used to optimise sensor placement based on binary optimisation techniques has also been developed as shown in [11], [12], and [13]. This is a convenient tool to optimise the visual sensors’ configurations when observing a target space such as a human activities space. Papers
III and IV describe the optimisation programme for 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 baseline length, visibility, camera movement, and depth reconstruction accuracy.

Since the IVAS is centred on the human being, the human factor should be taken into account when the system is designed. Essentially, a human factor can be considered as a human physical or cognitive property which restrains the design, [14]. To define the human activities space and practically implement it into the vision system, the modelling of a human activities space as a tetrahedron is presented in Papers III and IV.

In order to better understand human behaviour and perception, the proposed adaptive measurement method makes use of the fuzzily defined variable, FDV. The FDV approach enables the combination of qualitative and quantitative factors. The introduced method applies a neural network as a tool consisting of a learning function that integrates the qualitative factor into the IVAS. In the thesis, the image quality assessment using the FDV is presented. However, the FDV approach can be useful also when an IVAS attempts to learn from human behaviour and then generate a strategy for 3D reconstruction from the study of this behaviour.

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 types of methods, [15]. The quality index has been developed and used in different branches such as the food industry, [16], ecology, [17], and image processing, [18], [19], [20]. The image quality index has been proposed by Wang et al., [19], [20], where the image index correlates with the human visual perception. The image quality for the human visual perception can be introduced as the modelling problem of the FDV, as shown in Paper V.

1.1 Thesis objective

The objective of research that this thesis accounts for is to develop and validate the models that are a part of the Intelligent Vision Agent System. The implementations of these models enable the observation and interpretation of the environmental, dynamic changes caused by human activities. Furthermore, these models allow the integration of human subjective factors into the vision system.

To achieve this goal, the work has been focused on:

- Planning and control of the multi stereo visual sensor system;
- Measurement accuracy and quality;
- The human factor as a part of an intelligent vision system.

1.2 Thesis scope

The research project has focused on the human activities space, the vision control system, and the accuracy of the depth reconstruction methods. The human perception of image quality has been considered as well. The scope of three main issues of the IVAS presented in the thesis can be described as follows:

1. The depth reconstruction uncertainty is represented by the intervals between iso-disparity surfaces. The mathematical models for the general stereo pairs are also analysed. Furthermore, the depth reconstruction uncertainty analysis and
improvement by a dithering method is introduced. The dithering method is analysed for the standard parallel stereo cameras, and the parallel iso-disparity planes are used.

2. The human activities space is modelled as a tetrahedron. The vision sensor system helps to arrange the movement of a multi stereo visual sensor to acquire enough information for 3D reconstruction of the real scene. The sensors’ intrinsic and extrinsic characteristics are the parameters considered while optimising the configuration.

3. The image quality assessment can be carried out by means of the integration of quantitative and qualitative features and factors. The neural network is suitable for the integration of both qualitative factors and quantitative features.

1.3 Research methods

The thesis deals with theoretical and applied research related to the intelligent vision system. However, since the human factor is an important part of the system, the integration of quantitative and qualitative paradigms is required. The combination of the qualitative and quantitative methods may cause problems as quantitative methods might overwhelm qualitative ones. Hence, it is important to be aware of each component and to treat them all as parts of the puzzle, [15]. However, the main part of the thesis applies quantitative engineering research methods consisting of four stages deduced from constructive research, [21]:

- Problem identification: Asking the research question and addressing it with a possible hypothesis;
- Solution development: Exploring the theory and developing the models, algorithms, and tools;
- Solution implementation: Developing the practical methods that may be used to implement the theory and models in the system;
- Solution validation: Verifying the implementation results through simulations and real experiments.

Qualitative research relies on the ability to supply reasons for various aspects of human behaviour, [22], and the qualitative analysis applied in the thesis consists of four parts, data reduction, data collection, data display, and conclusions: drawing/verification, [23].

1.4 Thesis outline

The work presented in this thesis is based on the five papers reproduced in Part II. These papers contribute to the modelling of a human activities space, the planning of vision sensor system, and the adaptive measurement method for the image quality assessment.

The relations between the papers are illustrated by Figure 1.2. The common subject of all papers is the modelling and implementation of the IVAS. Papers I and II focus on the analysis and improvement of the depth reconstruction accuracy for a general target. The baseline length, sensor resolution, convergence angle, and the distance between the target and the camera are the factors that influence the accuracy the most. Furthermore, in Papers III and IV, the human activities space becomes the target space for the vision
sensor system, which is modelled by the tetrahedron. The integration of the human activities space and the vision sensor system is introduced in Papers III and IV. The optimal number of multiple stereo pairs, the selection of the multiple stereo pair baseline lengths, and the positions and poses applied to observe the human activities space are also described in these papers. Furthermore, Paper V presents the human perception which can be involved in the IVAS. Finally, the adaptive measurement method for image quality assessment is applied by the integration of qualitative factors and quantitative features.

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 the first chapter of Part I is to provide a brief overview of the relevant research areas and methods used in the thesis. In Chapter 2, the depth reconstruction method and accuracy analysis are presented. This chapter introduces the dithering algorithm for the depth reconstruction accuracy improvement. Chapter 3 describes the planning of multi stereo sensors used to monitor the human activities space through integer linear programming. Chapter 4 focuses on the adaptive measurement model applied in image quality assessment. A brief summary of the included papers, the conclusion of the thesis, and suggestions for future work are included in Chapter 5.

Figure 1.2. An overview of the relationship between the five papers.
2 Depth reconstruction method and accuracy

For a long time, people have wondered how we view the 3D world. During the 17th century, the question was routinely phrased as: *How does human depth perception work?* 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, the human face, 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 point, one needs at least two images. The necessary information regarding depth and the relations between objects can be found using those two images. Thus, it is possible to reconstruct a 3D model. Figure 2.1 shows the principle of using two images to reconstruct a point in a 3D space through a triangulation method. If we can observe the same points from two views, 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 is the point location in the space. The reconstruction from the two views is based on an epipolar geometry which describes the relation between the image points and the scene point. In order to obtain the 3D information, the image points’ information and the camera configurations are needed.

*Figure 2.1. The point in space can be reconstructed from two images by a triangulation method.*
The 3D reconstruction procedure essentially consists of the three following steps, [24]:

- Finding the corresponding image points for the same scene point;
- Obtaining the relative pose of the camera for the different views;
- Extracting the relation between the image points and their corresponding rays.

The relation between the image points and their corresponding rays is obtained from the pinhole camera model which is defined by the intrinsic and extrinsic parameters of the camera, [25].

The disparity, the quantity used in depth reconstruction, refers to the displacement of corresponding points on the left and right images along the corresponding epipolar lines for a common scene point. The critical problem of 3D reconstruction accuracy is thus to find the optimal sensor configuration.

The depth reconstruction accuracy depends on the system configuration which is defined by sensor resolution (pixel size), focal lengths, baseline length, and convergence angle. However, when determining the accuracy of a 3D reconstruction, the depth spatial quantisation is one of the most influential factors. This type of factor cannot be reduced even in more accurate measurements or configurations. 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, [26], [27]. Klarquist and Bovik presented a vergent active stereo vision system to recover the high resolution of depth by accumulating and integrating a multiresolution map of surface depth over multiple successive fixations, [28].

These considerations lead to the application of signal processing methods (e.g., dithering) when employing an active stereo camera system. The proposed method is shown in Paper II. The depth uncertainty analysis for a target space and the corresponding algorithm for optimising the number of stereo pairs and the stereo camera’s configurations are presented in Papers I and II.

### 2.1 The iso-disparity surfaces geometry model

The iso-disparity surfaces characterise the quantisation phenomena in stereo reconstruction, [3], [4]. The intervals between discrete iso-disparity surfaces represent the depth reconstruction uncertainty. The iso-disparity surfaces geometry models proposed in Paper I are valid for the convergence stereo pairs. There are two configurations for a stereo pair in common use: a convergence stereo pair and a parallel stereo pair.

A convergence stereo pair is the most general common configuration, where the optical axes cross at a fixation point. The simple mathematical model of iso-disparity surfaces for this configuration has been analysed in Paper I. The zero disparity circle is defined by the fixation point and the left and right camera optical centre position points. This circle is known as the Vieth-Müller circle and is a projection of the horopter, [29]. 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, while the
ellipses are cross sections of this cylinder on the optical axis plane. In order to define the ellipse position, shape, and orientation, we need to define the ellipse’s five degrees of freedom and its mathematical model. 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 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’ field of view. The geometry models show that the iso-disparity planes are parallel for a parallel stereo pair with the same focal length, and the iso-disparity planes converge to a straight line for the parallel stereo pair with a different focal length. The iso-disparity planes plots for these two configurations of the parallel stereo pair are shown in Paper I.

2.2 Depth reconstruction

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, small changes to the stereo convergence angle do not affect the depth accuracy very much, especially when the target is placed centrally.

The probability distribution functions of image 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 quantisation uncertainty interval of disparity equals the double image pixel size. The depth reconstruction quantisation uncertainty is the non-linear function of disparity and corresponds to the interval between the iso-disparity planes.

By adjusting the stereo pair’s profile, such as the baseline, the focal lengths, 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 3D reconstruction. Some signal processing methods can improve the accuracy. Dithering is one such possible method, and the usefulness of this method is explored in Paper II.

2.2.1 Depth reconstruction with dithering

In the proposed model of depth reconstruction, the left and right cameras are the quantisers. The quantiser input signals are the target point projection positions on the left and right image planes along the horizontal axis. The dither signals add noise to the signals prior to their quantisation in order to change the statistical properties of the quantisation, [7]. 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. An alternative is to shift the camera sensor, which means that the quantisation levels of the quantiser are changed. The proposed method is based on the movement of the camera sensor position.

The dither signal is a discrete one and is used to control the left and right cameras’ position. In Paper II, we have presented a two-stage discrete dither signal for each camera, which provides four images to calculate the depth of the target feature with an improved resolution and a reduced quantisation uncertainty.
The depth reconstruction uncertainty can be reduced by half when a dithering signal moves the new iso-disparity planes into the middle between the old disparity planes. Iso-disparity planes can be moved by increasing or decreasing the baseline. This can be accomplished by a single camera movement. To change the baseline length by placing the new iso-disparity plane in the middle between the old iso-disparity planes is also the optimal solution from a quantisation point of view. The analysis and example of a change in the baseline length are shown in Paper II. In this paper, it is shown that by aid of the proposed dithering method, the depth reconstruction uncertainty is reduced by half.

### 2.2.2 The implementation of depth reconstruction with dithering

From figure 2.2, it can be perceived that the discrete dither signals $d_{li}$ and $d_{ri}$ control the position of the left and right cameras. The dither signals are estimated by analysing 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 target point projections $x_{li}$ and $x_{ri}$ correspond to the $i$-th dither position of the left and right camera, respectively, and the quantised 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 following four steps:

1. The primary measurement of the target point depth is taken, where the target point is defined as the centre of the target object.
2. The dither signal is estimated and then generated by the baseline length change.
3. The secondary measurement and calculation of the new disparities are accomplished.
4. The final target point depth and its depth reconstruction quantisation uncertainty are calculated.

The dithering algorithm was verified through simulation and with the aid of a physical experiment in Paper II.

![Figure 2.2. 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 quantised signals for the left and right cameras, and $d_i$ is disparity.](image)
A human activities space, as a target space for 3D and depth reconstruction, provides constraints to the design and planning of the active stereo camera system. The sensor planning can be viewed as an extension of the well-known Art Galley Problem, AGP, [30]. 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, a similar calculation is required: the minimum number of stereo pair sensors needed to cover target space. The human activities space as a target space is defined here by a tetrahedron. In the field of active vision, there have been some studies on how the dynamical adjustment of the stereo baseline for one stereo pair may be used to improve the reconstruction accuracy, [31], [32]. However, there has been relatively little work on determining the optimum sensor configurations, [9].

This chapter gives an overview of the modelling of multi stereo sensor arrangements in the intelligent vision system. In Papers III and IV, we introduced camera constraints which focused on the visibility of the target. The accuracy constraint is based on the estimation of the depth reconstruction accuracy when the angles between the visual line of each camera and the baseline perpendicular are the same. The iso-disparity geometry model allows for a deepening of the analysis of the depth reconstruction accuracy. It analyses the entire camera Field of View, FoV. The accuracy constraint aids this process by dynamically adjusting the position, poses, and baseline lengths of multiple stereo pairs of cameras, thus acquiring the desired accuracy.

The planning algorithm proposed in Papers III and IV works in a 3D space. 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 minimal number of stereo pairs needed to cover a human activity space is determined with the aid of Integer Linear Programming, ILP, [11], [12], [33]. The 3D reconstruction accuracy, which is ensured by an accuracy constraint, can be further verified for a human activity space by a cubic reconstruction.

3.1 Constraints for the optimisation model
The constraints of the stereo view optimisation model can be determined from the environment, the camera, and the human properties. The environment, the camera, and the human properties significantly influence the system to identify and reconstruct the target. The details of each constraint are described in Papers III and IV.
The human activities space is modelled by a tetrahedron as shown in figure 3.1. The normal of each tetrahedron’s upper triangle gives the orientation of that surface. If the visibility angle, $\theta$, between the triangle normal and a line drawn from the centroid of the triangle to a specific camera position increases, this means that the image resolution decreases. In order to get a good image resolution, a visibility angle, $\theta$, of less than the maximum visibility angle, $\theta_{\text{max}}$, 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 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, $\phi$, must be less than the maximum angle, $\phi_{\text{max}}$.

In order to follow the movement of the target object, 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 which the system supports.

\[ \text{Dist}(\text{StereoPair}_{\text{next}}, \text{StereoPair}_{\text{current}}) \leq \text{Dis}_{\text{max}} \]  \hspace{1cm} (3.1)

The smaller number of potential next-view positions for the cameras restricted by (1) can simplify computation.

### 3.1.2 Constraints delivered from the stereo pair properties

The camera constraints are related to the camera FoV. The camera horizontal and vertical viewable angles, $\phi_h$, $\phi_v$, and a working distance, $r$, can be calculated from the
camera attributes (see the spherical coordinate systems shown in figure 3.2.) In order to keep the target object’s feature points within the camera FoV, the following constraints must be fulfilled:

\[
\begin{align*}
   l &\leq r \\
   \alpha_s - \phi_s / 2 &\leq \alpha_c + \phi_c / 2 \\
   \beta_s - \phi_s / 2 &\leq \beta_c + \phi_c / 2
\end{align*}
\]  

where \( l \) is the distance between the target position and the camera’s position; \( \alpha_s, \beta_s \) are the azimuth and the elevation of target, respectively; \( \alpha_c, \beta_c \) are the azimuth and the elevation of the camera’s pose, respectively.

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, \( B_{\text{max}} \).

### 3.1.3 The constraint of depth reconstruction accuracy

Depth reconstruction is one of major focuses of this research project. The depth reconstruction accuracy improvement can be adjusted by the baseline length, [9], [34]. This thesis suggests that a depth accuracy factor, \( AF \), is a function of the target convergence angle, \( \psi_t \), and the camera pose, \( \alpha_c \). In fact, it varies more significantly in respect to the target convergence angle than to the camera pose. Thus, the target convergence angle determines the depth accuracy factor. The accuracy constraint for a given point can be defined as:

\[ AF \leq AF_{\text{con}} \]  

---

**Figure 3.2. The spherical coordinate system and FoV of a camera where C is the camera position and the example target point is located at point T.**
where $AF_{con}$ is determined from the reconstruction accuracy requirements of the given application.

In order to further improve the reconstruction accuracy, the dithering algorithm for a parallel stereo pair presented in Paper II can be applied. The new iso-disparity surfaces that form after the dither signal has been added can be placed in the middle of the intervals of the previous iso-disparity surfaces. Thus, the depth reconstruction quantisation uncertainty may be reduced by half. The implementation of the parallel stereo pair is presented in Paper II. The implementation of the convergent stereo pair was extended from the parallel stereo pair.

3.2 Implementation of the camera planning with the integer linear programme

The stereo pair placement planning consists of three stages:

- Firstly, with the aid of the greedy algorithm, we find potential stereo pairs that satisfy the stereo constraints from all potential cameras’ positions and poses, as presented in Paper III.
- Secondly, integer linear programming is applied to minimise the total stereo pairs subjected to the visibility and baseline length constraints, depth accuracy constraints, and 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 accuracy can be verified by a cubic reconstruction.

The 3D simulations for human body and activities space coverage by stereo pairs are presented in Papers III and IV.
An adaptive measurement method

For computer monitoring to be effective and useful, a measurement of the human mood or health needs to be introduced into the IVAS. However, these human characteristics vary from person to person and depend on many factors. Such quantities cannot be precisely defined and do not possess any standards. To model and measure them, one must use methods which have to adapt to each individual and his/her personal characteristics. Such quantities are defined as “fuzzy”, and they are discussed in Paper V with the aid of the Fuzzily Defined Variable, FDV. An adaptive method for the measurement of a FDV which can be applied for the purpose described above is introduced in this chapter.

4.1 An adaptive measurement method and its implementation

The FDV often consists of both quantitative and qualitative factors, both of which are of different importance for different targets or users. The FDV attributes 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. Since the FDV often depends on both quantitative and qualitative factors, it is difficult to express it in only quantitative terms, [35]. 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 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 could be generalised for many different kinds of purposes. The measurement method can be adjusted to changeable parameters. The method uses:

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.
Figure 4.1 illustrates the modelling of a quality index that applies an adaptive measurement method. The initial quality index model is established by experts in the field. The set of quantitative features to be included in the measurement of the quality index and the features’ initial weights, \([\alpha]\), are based on the measurement uncertainty and relevance of each feature. Then, the adaptive measurement method applies a training process to integrate the relationship between the value of the quantitative features and the subjective human assessments regarding quality.

Since the quality index of the product, service, or condition is used for different purposes, the human assessment can differ radically. In these cases, a group classification method is useful. The judges are grouped according to different factors that may determine how they subjectively assess the quality. Such factors include the purpose, age, gender, personality, background, etc., of a particular product or judge. The group classification method is based on Principal Components Analysis, PCA. The applied group classification procedure is as follows before applying the adaptive quality model:

- 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, is 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.

![Figure 4.1. A block diagram illustrating the quality model. Ellipses denote representations of information, and rectangles denote process transformations from one representation into another.](image-url)
A highly useful implementation tool that can be applied on the proposed adaptive measurement method, used to determine the quality index is the Neural Network, NN. During the training stage, two input data – the 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, [36]. Then, the trained quality model estimates the discrete QI of the product/service/condition 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 \([\alpha]\).
2. Group classification, by finding the correlation between human assessment and qualitative factors.
3. Training stage for self-organising the NN input layers according to classified groups and the estimation of the weights of NN.
4. Validation stage, to get the discrete QI.

Figure 4.2 shows the validation stage of the adaptive system using the NN. 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.

4.2 The adaptive method for image quality measurement

When assessing image quality, several multidimensional aspects need to be considered. There are different image quality indices, depending on the application area. A new image quality index is proposed by Wang et al., [19], [20]. 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. The image quality index can be useful for the IVAS as the image processing algorithm is selected and the visual sensors’ positions and poses are chosen. Image quality is influenced by quantitative features such as basic properties, naturalness, and colourfulness. The initial weight of each quantity is estimated by experts based on the quantitative factors’ measurement uncertainty and cost, as well as their relevance. However, the human quality assessment of image quality depends also on many qualitative factors such as personal background, physical environment, usefulness, tools, and pattern representation, which are related both to the target and the human being.
The simplified image quality model has been tested by an estimation of the QI of greyscale images. During this test, the applied NN consisted of two stages: a training stage and a validation stage.

In the first step of the procedure, we chose to classify the images according to different groups of people. The groups’ 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 principal components resulting from the application of the PCA. Next, the mean values of the reconstructed grades were taken for each image within each group. However, the results from Paper V show that the different groups provide compatible quality assessments. Therefore, it was considered unnecessary to distinguish between the different groups of people participating in the study.

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, [36]. Afterwards, the Back-Propagation Neural Network approach was applied. Three types of quantitative features were used during the training stage: structure distortion ratio, along with the two basic properties luminance distortion ratio and contrast distortion ratio. These measurements of the intensity data of the pattern and test images were normalised, [19], [20]. The model was trained on the same image but with different disturbances.

The validation stage occurred after the training stage of the adaptive system with the help of the NN. The QI developed by the Neural Network and the ranking produced by the human judgments matched each other very well. Based on the result from Paper V, one can conclude that the model recognises different kinds of disturbance.

4.3 Conclusions and further development
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 the quality of a particular product, service, or condition. 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 adaptive quality image model 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.

This model can be further developed by looking at the adaptive methods that may be used to determine a human mood/health index from the visual observation of human habits. The human mood/health index should aid the dynamic estimation of human health in a more suitable and objective way. The system may use the NN as the tool to generate a human health index that may be used to predict human health. This prediction can be useful for pre-diagnostics. The human health approach will be one of the main applications of an IVAS system.
5.1 Summary of contributions

This chapter gives a brief summary of the five papers included in the thesis. The papers are presented in logical order, as described in Figure 1.2. The summary describes how the studies were performed, the results obtained, and the conclusions drawn.

5.1.1 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 iso-disparity surfaces, and the depth reconstruction accuracy is defined by the intervals between the iso-disparity planes. A mathematical geometry model is used to analyse the iso-disparity surface. This model can be used to dynamically adjust the positions, poses, 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 the function of the baseline length, focal lengths, and 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 to the convergence angle. Small changes in the stereo convergence angle do not affect the depth accuracy overly much, especially 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.

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 systems configuration and target properties.

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

The depth reconstruction uncertainty is 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 depth reconstruction uncertainty mathematical model used for analysis of the reconstruction uncertainty is presented in the paper. A dithering algorithm is
implemented to reduce the depth reconstruction uncertainty, and its 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 an accepted 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 of 49.7%. The physical experiment shows an improvement of the depth reconstruction accuracy of 36.2%. The results furthermore revealed that the target depth reconstruction uncertainty is reduced by half by the proposed algorithm. The differences in results from the simulation and the physical experiment can be attributed to other factors that influence the measurement.

5.1.3 Paper III - 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, object, and stereo pair constraints into the integer linear programming.

The camera’s 3D field of view is modelled by spherical coordinates, which speeds up computations. 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 a 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’ poses 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 poses according to the target location and required constrains. The simulations proved that a set of two pairs is sufficient to observe the human activities space modelled as a tetrahedron, under the condition that in each position all upward triangle surfaces are visible to at least one stereo pair.

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

In this paper, the key factors which affect the accuracy of 3D reconstruction are analysed. The paper argues that the convergence angle and 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 pose. The depth accuracy constraint guarantees 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 baseline length control 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 a two-stage camera sampling is implemented. The two-stage camera position sampling allows for flexible adjustment of the position ranges and intervals and thus speeds up computation.

5.1.5 Paper V - 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 NN is used. With the help of NN, 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 QI estimated by the adaptive system and the human quantitative assessments matched each other very well.

5.2 Conclusions

This thesis focuses on depth reconstruction, the planning of multiple stereo sensors that can be used to monitor a human activities space, and an adaptive model of quality assessment, all of which are important elements of an IVAS. The research work can be summed up through a discussion of two key issues: the uncertainty analysis and the handling of the human factor by vision systems.

The depth reconstruction uncertainty is illustrated by the iso-disparity surfaces model. The model facilitates a quantitative analysis of the depth reconstruction uncertainty and can also optimise this uncertainty by adjusting the multiple stereo pairs’ positions, poses, and baselines. The depth reconstruction uncertainty map is calculated by means of the iso-disparity surface model. The iso-disparity surfaces model is also a robust model that can be used when dynamically controlling the stereo pair’s baseline and the camera’s corresponding positions and poses in order to observe a moving target. Furthermore, since a discrete sensor causes the depth spatial quantisation uncertainty, which is one of the most significant factors influencing the depth reconstruction accuracy, the dithering algorithm is a suitable method for reducing the depth reconstruction uncertainty. The proposed algorithm assures high precision depth reconstruction from a few images taken by low-resolution sensors. The depth reconstruction uncertainty can be reduced by half by the dithering approach compared to the direct triangulation method. The dither signal is analysed by means of the iso-disparity planes and then generated by controlling the stereo pair’s 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 the camera movement.
The suggested multiple stereo vision sensor planning guarantees that the target is observed under efficient visibility and the required depth reconstruction accuracy. The constraints of the target space, the stereo pairs’ properties, and the reconstruction accuracy have been explored in the research conducted as a part of this thesis. Furthermore, the implementation of the integer linear programming model minimises the amount of stereo pairs necessary to cover the entire target space under efficient visibility and the required depth reconstruction accuracy. The depth reconstruction accuracy is ensured by the accuracy constraint and further verified through the use of a cubic model. Considering the impact of the human factor on the vision system, the human activities space is modelled by a tetrahedron, which gives a convenient way to extract the orientation of each surface and guarantee good observability.

In addition to this, the introduced adaptive measurement method which makes use of the FDV integrates the human factor into the measurement uncertainty estimation for image quality assessment and human habits prediction. The model is very useful for cases where there is a lack of a clear definition of the quantity, e.g., the image quality related to individual human perception. The adaptive model proposed in the thesis has been successfully implemented for image quality assignments where both quantitative and qualitative factors must be considered. As an implementation tool, the Neural Network has been used.

5.3 Future research

There are many new research questions and possible research problems that can be highlighted as a result of the research presented in this thesis. For example, it would be both interesting and important to look into the improvement of the human geometry model. It would be possible to use the model proposed in this thesis to distinguish between two of the upper triangles of the tetrahedron which represent the human face side. It would also be interesting to study how the movement of the camera can be planned by considering the target occlusions and zooming for the most informative parts.

In particular, it would be useful to analyse the two dimensional image quantisation uncertainty and suitable dithering methods. Such research could be useful not only for 3D reconstruction but would also aid the study of stereo matching algorithms. The dithering algorithm can be used to improve the information of particular image points and thus reduce the matching uncertainty between the corresponding stereo points.

Interesting 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. The blurred part should be transformed after the application of the dithering algorithm. In this case, it could be possible to find more depth information from the range of the blurred part.

Furthermore, during our research, it became apparent that the camera’s micro movement caused by the dither signal affected not only the projection position of the target point but also changed the projection image pixel intensity. This suggests that depth reconstruction using the dithering algorithm in relation to the projection image pixel intensity changes needs to be developed further to achieve more accurate depth reconstruction.
The idea presented in Paper $V$ is a promising starting point for introducing the quantitative human factor into the vision system in the future. This could lead to potential dynamic learning of human habits or the human health/mood status, then applying 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 of quality assessment, the possibility of an intelligent television can be explored. The IVAS can recognise, identify, and learn the human preferences regarding the television screen’s colour, contrast, saturation, etc. According to human individual preferences, the system can then make decisions regarding the support of the peripherals to automatically control the television’s settings and surrounding environment such as lights, curtains, etc. Another application example is how the adaptive system may find the proper stereo pair positions, poses, and image processing algorithms to improve the image quality for health/mood monitoring and diagnostics applications. The adaptive measurement method of quality can be applied to the 3D reconstruction accuracy strategy which can then be adjusted according to both human and environmental factors.

The work presented in the thesis constitutes a first step towards advanced IVAS design. There are many scientific problems and questions which have to be explored and further developed.
References


Part II
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:
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 Field of View, 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:

\[
z(x,n) = \frac{f_L - f_R}{n \Delta D} x + \frac{B}{2 n \Delta D} (f_L + f_R)
\]

where \( B \) is baseline length, \( n \) is integer number , \( \Delta D \) 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:

$$x = \frac{B}{2} \left( \frac{f_L}{f_R} + \frac{f_R}{f_L} \right), \quad z = 0 \quad \text{for} \quad f_L \neq f_R$$

(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 $\Delta D = 0.04$ 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 $\alpha_c$, where $\alpha_{cL} = \alpha_{cR} = \alpha_c$ 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, $(\psi_{L0}, \psi_{R0})$, 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 $M_{\infty}$. $M_{\infty}$ 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 $\alpha_{cL0} = \alpha_{cR0} = \alpha_c$, 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 $M_{\infty}$. 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.
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 \( \alpha_{cL0} = \alpha_{cR0} = \alpha_c \), describes a cylinder, the ellipses being cross sections of this cylinder on the \( xz \)-plane with centres in \( 0_e(x_0(n), z_0(n)) \):

\[
\frac{(x - x_0(n))^2}{a^2} + \frac{(z - z_0(n))^2}{b^2} = 1
\]  

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, \( f_L = 3.5 \text{ cm}, f_R = 3.0 \text{ cm} \) for left and right camera respectively. The convergence point is \((-195 \text{ cm}, 0)\) on \( xz \)-plane. (b) Cameras with the same focal length of 3.25 cm. (c) Cameras with different focal lengths \( f_L = 3.0 \text{ cm}, f_R = 3.5 \text{ cm} \) for left and right cameras respectively, the convergence point is \((195 \text{ cm}, 0)\) on \( xz \)-plane.

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 \( x_0 = 0 \) and \( z_0 = b - B \tan \alpha_c / 2 \):

\[
\frac{x^2}{a^2} + \frac{\left( z - \left( b - \frac{B}{2} \tan \alpha_c \right) \right)^2}{b^2} = 1
\]  

(4)

where \( B \) is the baseline length and \( \alpha_c \) 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 \( \alpha_c \) and is described as:

\[
b = \frac{B}{\sin 2\alpha_c - \frac{n \Delta D}{f} \cos^2 \alpha_c}
\]  

(5)

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

\[
\left( \frac{b}{a} \right)^2 = \frac{1 + \frac{n \Delta D}{2 f} \tan \alpha_c}{\tan \alpha_c - \frac{n \Delta D}{2 f}} = \frac{\tan \alpha_c}{\tan \psi_c}
\]  

(6)

where \( \psi_c = \psi_{L,0} = \psi_{R,0} \).

The result of the synthetic stereo pair simulation with a baseline length 50 cm, focal lengths 2.5 cm and disparity resolution \( \Delta D = 0.04 \) 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 in Fig. 3(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, \( \alpha_c = 4^\circ \), the baseline length \( B = 40 \) cm, the focal lengths \( f = 2.5 \) cm and disparity resolution \( \Delta D = 0.04 \) cm (a) perspective view, and (b) top view.](image-url)
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 $M_f$. 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 $\Delta D$ 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 $M_f$.

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, $\alpha_c=4^\circ$ and the FoV is approximately $54^\circ$. 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.](image-url)
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.

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. 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.
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 fulfill 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

![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.](image)

![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.](image)
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


Depth Reconstruction Uncertainty Analysis and Improvement –
The Dithering Approach

Authors:
Jiandan Chen, Siamak Khatibi and Wlodek Kulesza

Reformatted version of paper submitted to:
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 the 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 the depth reconstruction error statistics. The physical experiment proves 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. 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, [1-2]. When determining the accuracy of 3D reconstruction, the 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 is then reduced, and also because of the restricted sensitivity of the sensor itself. The selection of an optimal sensor pixel size is discussed in [3].
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, [4-5]. The quantization uncertainty of depth reconstruction is discussed in [6]. However, there has been little work on reducing the depth reconstruction uncertainty based on the combination of low-resolution images. In this paper, the dithering algorithm combines the dithering and the 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 is introduced in [7-10], and proves that the system resolution decreases 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, [11]. In relation to this, Yang [12] 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 [13] also use dithering to locate a subpixel edge in a binary image. Klarquist and Bovik presented a vergent active stereo vision system to recover the high resolution of depth by accumulating and integrating a multiresolution map of surface depth over multiple successive fixations, [14].

In general, the camera sensor consists of uniformly distributed two dimensional arrays. 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 [15] 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. [16] 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.
2. Problem statement and main contribution

2.1 Problem statement
At least two images are needed to obtain a depth perception 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 described in pixel. This leads to a depth reconstruction quantization uncertainty in the representation of the spatial position of a scene. The depth reconstruction uncertainty is proportional to the pixel size, [6], and becomes a serious issue when the pixel size is large compared to the allowable tolerance in 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 to 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.

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 \( \Delta \) (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 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:

\[
Z_{nt} = \frac{Bf}{|x_{Ql} - x_{Qr}|} = \frac{Bf}{n_t \Delta}
\]  

(1)

where \(B\) is the baseline length, \(n_t\) is a target point disparity which is an integer number, calculated from \(n_t = |x_{Ql} - x_{Qr}|\), and \(\Delta\) is a pixel resolution. For a parallel stereo pair, 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.

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

\[
x_l = \frac{f}{Z_t} \left( X_t + \frac{B}{2} \right) \quad x_r = \frac{f}{Z_t} \left( X_t - \frac{B}{2} \right)
\]  

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 the further analysis focuses on
general quantization formulae, \( x_l \) and \( x_r \) are denoted as \( x_c \). From this follows that the
corresponding image horizontal quantization uncertainty, \( q_h \), can be described as:

\[
q_h(x_c) = \frac{\Delta}{2} - \Delta \left\{ \frac{x_c}{\Delta} + 0.5 \right\}
\]  

where the symbol \( \left\{ \cdot \right\} \) denotes the fractional part. The sawtooth-shaped function (3) is a
periodical function of \( x_c \) with the period, \( \Delta \), and can be expressed as a Fourier series,
[17]:

\[
q_h(x_c) = \sum_{k=-\infty}^{\infty} \frac{\Delta}{\pi k} \sin \left( \frac{\pi k x_c}{\Delta} \right)
\]  

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 the quantization uncertainties of the left
and right images are uncorrelated [18], the probability density functions, PDFs, of the
image horizontal quantization uncertainties, \( p_{hl} \) and \( p_{hr} \), for the left and right images
respectively are:

\[
p_{hl}(q_{hl}) = \frac{1}{\Delta} \text{ where } -\frac{\Delta}{2} \leq q_{hl} \leq \frac{\Delta}{2} \quad \text{and} \quad p_{hr}(q_{hr}) = \frac{1}{\Delta} \text{ where } -\frac{\Delta}{2} \leq q_{hr} \leq \frac{\Delta}{2}
\]  

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:

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

where \( \otimes \) denotes convolution. The result of the convolution of two rectangular
distributions is a triangular one. The quantization uncertainty interval of disparity, \( q_d \), is
2\( \Delta \) 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 image coordinates on the x-axis for a common scene point is called disparity and is presented in (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 $\Delta$, [6]. By using (1), we can estimate the target point’s discrete depth, $Z_{mt}$, 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:

$$q_{Zt} = \frac{Bf}{n_t(n_t + 1)\Delta} \quad (7)$$

### 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, [8]. 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 by which to add 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 equispaced and equiprobable impulses which numbers are greater than or equal to two. In this paper, we propose a two-stage discrete dither signal for each camera, which means four images to calculate a depth of the target feature with 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:

$$f(D) = \frac{1}{2} \left[ \delta\left(D - \frac{D_a}{2}\right) + \delta\left(D + \frac{D_a}{2}\right) \right] \quad (8)$$

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

The characteristic function, $\Phi$, of the PDF of the dither signal is:

$$\Phi(\omega) = \frac{1}{2} \left( e^{j\omega \frac{D_a}{2}} + e^{-j\omega \frac{D_a}{2}} \right) = \cos\left(\frac{D_a}{2} \omega\right) \quad (9)$$

where $\omega$ is the frequency angle of the signal.

The mean value of the image horizontal quantization uncertainty, $q_{h}$, with the discrete dither signal $D$ is:
where $E(\cdot)$ is the expected value operator. By using (4) and (9), we can get, [16]:

\[
\bar{q}_h(q_h) = \sum_{k=1}^{\infty} \frac{\Delta}{\pi k} (-1)^{k+1} \Phi \left( \frac{2\pi k}{\Delta} \right) \sin \left( 2\pi \frac{k}{\Delta} q_h \right)
\]

(11)

where $x_c$ can be substituted by $-q_h$ due to its periodicity with the period, $\Delta$.

From (9) and (11) we can get:

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

\[
= \sum_{k=1}^{\infty} \frac{\Delta(-1)^{k+1}}{2\pi k} \left[ \sin \frac{\pi k}{\Delta} (2q_h - D_a) + \sin \frac{\pi k}{\Delta} (2q_h + D_a) \right]
\]

(12)

where $q_h$ varies between $-0.5\Delta$ and $0.5\Delta$.

When $\Delta$ is normalized into 1 pixel, Fig. 2 shows the plot of the function, $q_h$, versus $q_h$ with a range from $-0.5\Delta$ to $0.5\Delta$ for five different amplitude values of the discrete dither signal, $D_a=\{0.1\Delta, 0.4\Delta, 0.5\Delta, 0.6\Delta, 0.9\Delta\}$.

From Fig. 2, it is clear that the mean image horizontal quantization uncertainty interval of an image with the discrete dither signal amplitude $D_a = 0.5\Delta$ has a minimum

![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 $D_a$.](image)
interval of 0.5Δ and varies in the range of -0.25Δ to 0.25Δ. Applying the result to (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 (6) that the interval of the disparity quantization uncertainty, \( q_d \), without dithering is 2Δs in the range of [-Δ 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 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 the focal length \( f \) and pixel size \( \Delta \).

In the 3D space, the depth reconstruction uncertainty is represented by the intervals between the iso-disparity planes according to (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:

\[
D_a(n_t) = \frac{n_t \Delta}{2(n_t + 1)}
\]  

(13)

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

\[
\Delta B_t = \frac{B}{2(n_t + 1)} = \frac{B \cdot D_a(n_t)}{n_t \cdot \Delta}
\]  

(14)

In Fig. 3(a), 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 length \( \Delta = 12.9 \mu \text{m} \) and both cameras have the same focal length \( f = 25 \text{ mm} \). For instance, if the target point’s disparity \( n_t = 49 \), the measure depth is \( Z_{49} = 1997 \text{ mm} \). To shift the iso-disparity plane further away from the baseline by a half of the depth reconstruction quantization uncertainty, \( q_Z / 2 \), the baseline length has to increase by 0.51 mm. In Fig. 3.(a), green dashed lines show the new iso-disparity planes for the stereo pair with a baseline length of \( B’ = B + \Delta B_t = 50.51 \text{ mm} \). The zoomed iso-disparity planes with \( Z_n \) ranging from 1600 mm to 2400 mm are shown in Fig.3. (b). 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 (13), 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, \( \bar{q}_x = 0.5\Delta \). The change of the baseline length according to (14) 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.
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 $0.5 \Delta$, for a target point $T_0$ with the disparity, $n$. 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 $\Delta D_a$, and define it as the difference between the projections of the target point $T_0$ and any point $T$ on the stereo images with disparities, $n$ and $n'$ respectively. It can be formulated as:

$$\Delta D_a(n) = |D_a(n) - D_a(n')| = \frac{n - n'}{2(n + 1)} \cdot \Delta$$

Assuming 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 (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, $\Delta D_a$, is $0.1 \Delta$. The points in the accepted space with the disparity, $n_i$, where $\pm 0.3$ is the mean image horizontal quantization uncertainty tolerance $\Delta \bar{q}_h$, can be gotten from:

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

where the symbol $\lfloor \cdot \rfloor$ denotes the integer part. Furthermore, due to (13), the target point disparity $n_i$ 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 disparity, $n$, for three different target points with the disparities, $n_i = 20, 49$ and 176 respectively, the baseline length $B = 50$ mm, and the focal

**Fig. 3.** (a) The iso-disparity planes for parallel cameras where $\Delta = 12.9 \mu m$, and $f = 25$ mm; the red solid lines for $B = 50$ mm, 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.
length $f=25$ mm. Table I shows the disparity ranges, $n_{0.3}(n_t)$, which correspond to the depth ranges $\Delta Z_{0.3}(n_t)$ for a target point with the 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

<table>
<thead>
<tr>
<th>$n_t$</th>
<th>Disparity range ($n_{0.3}(n_t)$)</th>
<th>Depth range ($\Delta Z_{0.3}(n_t)$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>15 - 24</td>
<td>6460 mm - 4037 mm</td>
</tr>
<tr>
<td>49</td>
<td>39 - 59</td>
<td>2485 mm - 1642 mm</td>
</tr>
<tr>
<td>176</td>
<td>140 - 211</td>
<td>692 mm – 459 mm</td>
</tr>
</tbody>
</table>

4. Implementation

From Fig. 5, it can be perceived 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 $\Delta=12.9 \mu m$, $B=50$ mm and $f=25$ mm.
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 (1) considering the stereo pair cameras’ initial positions.

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

According to (14), 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, 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 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 $\mathbf{x}_{Ql}$ and $\mathbf{x}_{Qr}$:

$$
\mathbf{d} = \mathbf{x}_{Ql}^T \cdot (-\mathbf{x}_{Qr})
$$

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

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

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

$$
\text{AVG.}
$$

where $D_{li}$ and $D_{ri}$ are the dither signals for the left and right cameras, $\mathbf{x}_{Qli}$ and $\mathbf{x}_{Qri}$ are the quantized signals for the left and right cameras, and $d_i$ is disparity.
four pairs of images, which can be described as:

\[ Z_i = \frac{1}{4} \sum_{j=1}^{2} \sum_{t=1}^{2} \frac{fB_{i,j}}{d_{i,j}} \]  
(18)

with

\[ B = \begin{bmatrix} B & B + \Delta B_i \\ B - \Delta B_i & B \end{bmatrix} \]

where \( B \) is the baseline for the stereo pair camera with its initial position, while \( \Delta B_i \) 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 cases 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 [19]. The simulation experiment was performed in MATLAB 7.0. The Epipolar Geometry Toolbox, [20], was used to project and transform the object position in 3D space. In both cases studies, the cameras’ optical axes were in the \( XZ \)-plane, the initial stereo pair baseline length was 50 mm, focal length was 25 mm and pixel length \( \Delta = 12.9 \mu m \).

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.6 (a). 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.6 (b). 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 (14), we can determine that the left and right camera movement distance \( \Delta B = 0.51 \) mm. All points in the cubic space are in a disparity range of 46 to 53, and according to (15) their maximum deviation projection displacement \( \Delta D_{\alpha} \) is 0.04\( \Delta \), which is inside the accepted space with a mean disparity quantization uncertainty of \( \pm 0.6 \Delta \). In Fig. 6(c), the black points form the new iso-disparity planes reconstructed by 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 simulation points by both the direct triangular method and the dithering algorithm are shown in Fig. 7. The histograms prove 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, $\sigma$, 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.*
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 between two targets, \( \Delta Z_L \). The physical experiment was designed to measure a line differential depth of \( \Delta Z_L \) 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 original space coordinate is located at the center of the lens and the linear translation stage table is parallel with the \( X \)-axis. A ruler functions as the target object and it is transported by the linear

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

Fig. 8. The photo and schematic diagram of objects and configurations in the physical experiment.
translation stage table along the X-axis with the resolution of 1/80 mm. The ruler is located about 550 mm away from the camera lens center and has the angle, \( \alpha \), 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, \( \alpha \), 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 the base line length of 50 mm. 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 0.03\( \Delta \), 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 (14), the baseline should increase (decrease) by 0.14 mm. After shifting the ruler we took the other left and right images. The \( \Delta Z_L \) 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 triangulation method and the dithering algorithm. It 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 \( \Delta Z_L \) showing errors by the triangulation method and dithering algorithm for the real target lines.

<table>
<thead>
<tr>
<th>Angles ( \alpha: ) Line</th>
<th>Exact ( \Delta Z_L ) [mm]</th>
<th>The direct triangulation method</th>
<th>The dithering algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \Delta Z_L ) [mm]</td>
<td>Reconstruction error [mm]</td>
</tr>
<tr>
<td>35.6 (^\circ)</td>
<td>1</td>
<td>46.5</td>
<td>43.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>46.5</td>
<td>41.9</td>
</tr>
<tr>
<td>19.7(^\circ)</td>
<td>1</td>
<td>27.0</td>
<td>21.9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>27.0</td>
<td>21.3</td>
</tr>
<tr>
<td>0(^\circ)</td>
<td>1</td>
<td>0</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>6.4</td>
</tr>
</tbody>
</table>

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 the 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 model and control the movement of the cameras used to get these different images.
Analysis of the depth reconstruction uncertainty distribution is based on the analysis of image disparities and the iso-disparity approach. The interval between the iso-disparity planes represents the depth reconstruction quantization uncertainty and the interval is also the non-linear function of the target disparity. The probability density function of the disparity quantization uncertainty is a triangular distribution as a result of convolution of two rectangular distributions of the probability density functions of the left and right image horizontal quantization uncertainty. The quantization uncertainty interval of disparity equals the double pixel size and varies between $-\Delta$ and $+\Delta$.

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 the 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 presented model can be also applied to define an accepted 3D reconstruction space with defined accuracy. This extends a more realistic target point into this space.

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 algorithm consists of four steps. First, the primary measurement of the target point depth is taken. Then, the dither signal is estimated and generated by the baseline length change. Finally, after the secondary measurement and calculation of the
new disparities, the final target point depth and its depth reconstruction quantization
uncertainty are calculated.

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 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 of accuracy.

Further research is needed to explore this issue. In particular, it would be useful to
analyze the two dimensional image quantization uncertainty and its corresponding
dithering methods. 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.

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


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:
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 Intelligent Vision Agent System, 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.
1.1 Related works

The sensor planning can be viewed as an extension to the well-known Art Gallery Problem, 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 Field of View, 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 Integer Linear Programming, 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.

*Image Resolution*, 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:

\[
\phi_h = 2 \arctan \frac{S_h}{2f}, \quad \phi_v = 2 \arctan \frac{S_v}{2f}
\]

where \(\phi_h\) and \(\phi_v\) 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 \(\alpha_c\) and elevation \(\beta_c\).
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:

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

(2)

The azimuth $\alpha_o$ and elevation $\beta_o$ of target object with respect to camera position are given by

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

(3)

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

$$l \leq r \quad \text{and}$$

$$\alpha_c - \phi_h / 2 \leq \alpha_o \leq \alpha_c + \phi_h / 2,$$

$$\beta_c - \phi_i / 2 \leq \beta_o \leq \beta_c + \phi_i / 2$$

(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 $T_{v1,2,3,4}$, as in Figure 3. The three upward triangles are required to be covered by cameras’ FoVs. The normal of each
triangle gives the orientation of the surface. If the visibility angle $\theta$, 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 $\theta$ less than the maximum visibility angle $\theta_{\text{max}}$ is required:

$$\theta \leq \theta_{\text{max}}$$ (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 $\varphi_{\text{max}}$ is also required and constrained as:

$$\varphi \leq \varphi_{\text{max}}$$ (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):

$$Z = \frac{Bf}{dx}$$ (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; $\theta$ - the visibility angle between the triangle normal $\vec{c}$ and a line from the centroid of the triangle to the camera position; $\varphi$ - 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:

\[ B \leq B_{\text{max}} \]  

(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, \( \text{Camera}^k_{C_n, \alpha_n} \), 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, \( \text{StereoPair}_i \), 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 \( \text{StereoVis}_{ij} \) is defined by:

\[
\text{StereoVis}_{ij} = \begin{cases} 
1 & \text{if a StereoPair}_i \text{ covers triangle } j \text{ of target object model} \\
0 & \text{otherwise}
\end{cases} 
\]

(9)

which indicates each triangle \( j \) as the row \( j \) to be covered by the stereo pair \( i \) in the column \( i \), and \( 1 \leq i \leq K_s \), 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:

\[
\min \sum_{i=1}^{K_s} S_i 
\]

subject to
\[ \sum_{i=1}^{K_s} S_i \times \text{Stereovis}_{j,i} \geq 1, \quad \text{for } j = 1, 2, 3 \]  

(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:

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

(12)

for \( i = 1, 2, \ldots, K_s \), and \( p = 1, 2, \ldots, K_s \).

One more constraint is added into the model:

\[ \sum_{i=1}^{K_s} S_i \times c_{p,i} \leq 1, \quad \text{for } p = 1, 2, \cdots, K_s \]  

(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 \textit{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 \( \phi_h, \phi_v \) of 60° and has a working distance \( r \) of 7 m. The maximum visibility angle \( \theta_{\text{max}} \) and the angle \( \phi_{\text{max}} \) are taken to be 70° and 10° respectively. The maximum stereo baseline length \( B_{\text{max}} \) 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 $\alpha_c$ and elevation $\beta_c$ 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.

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.
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.

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-


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


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:
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 Field of View, 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, $\phi_h$, $\phi_v$, 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:

\[
\frac{l}{r} \leq r \quad \text{and} \\
\alpha_c - \phi_h/2 \leq \alpha_c \leq \alpha_c + \phi_h/2, \\
\beta_c - \phi_v/2 \leq \beta_c \leq \beta_c + \phi_v/2.
\]  

(1)

where: $l$ is the distance between the target position and camera’s position; $\alpha_c$, $\beta_c$ are respectively the azimuth and elevation of target; $\alpha_c$, $\beta_c$ 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, $\theta$, 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, $\theta$, of less than the maximum visibility angle, $\theta_{\text{max}}$, is required:

$$\theta \leq \theta_{\text{max}}$$  \hspace{1cm} (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, $\varphi$, of less than the maximum $\varphi_{\text{max}}$ is required and is constrained as:

$$\varphi \leq \varphi_{\text{max}}$$  \hspace{1cm} (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, $B_{\text{max}}$:

$$B \leq B_{\text{max}}$$  \hspace{1cm} (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, $(\alpha_{\ell}, \alpha_{r})$, are the angles of each camera rotated inwards from the parallel to achieve convergence. The target convergence angles, $(\psi_l, \psi_r)$, are the angles between the visual lines of each camera and the baseline perpendicular. From Fig. 1, simplifying: $\psi_l = \psi_r = \psi$, and $\alpha_c = \alpha_{\ell} = \alpha_{r}$, we obtain:

$$Z = \frac{B}{2 \tan \psi}$$  \hspace{1cm} (5)

where: $B$ is a baseline length and $Z$ is a target distance.

The equation (5) can be written as:

$$Z = \frac{B}{2} \left( \frac{1}{\tan \alpha_c - \tan(\alpha_c - \psi)} + \frac{\tan \alpha_c \cdot \tan(\alpha_c - \psi)}{\tan \alpha_c - \tan(\alpha_c - \psi)} \right)$$  \hspace{1cm} (6)

In the case of parallel stereo or with the target close to the fixation point, the $\alpha_c$ or $(\alpha_c-\psi)$ varies by a small amount, and the equation (6) can be further simplified. The resolution of the target convergence angle, $\psi$, is related to a single pixel, $p$, in the image, thus the relative depth error can be written as:

$$\frac{\Delta Z}{Z} \approx \left| \frac{\cos \alpha_c}{\sin \psi \cos(\alpha_c - \psi)} \right| \frac{p}{f} = AF \cdot \frac{p}{f}$$  \hspace{1cm} (7)

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, $\psi$, than to the camera’s pose, $\alpha_c$, therefore the target convergence angle determines the depth accuracy factor. The accuracy constraint for a given $p$ can be defined as:

$$AF \leq AF_{\text{con}}$$  \hspace{1cm} (8)

where: $AF_{\text{con}}$ 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:

$$\text{Dist}(\text{StereoPair}_{\text{next}}, \text{StereoPair}_{\text{current}}) \leq \text{Dis}_{\text{max}}$$  \hspace{1cm} (9)

where: $\text{Dis}_{\text{max}}$ 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, $AF_{\text{con}}$, 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 perspective projection matrix, 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, \( \Delta_{rec} \), for a single pixel error along a horizontal direction in the rectified image, has same value as the depth error and is given by:

\[
\Delta_{rec} = \sqrt{\frac{1}{8} \sum_{i=1}^{8} (\hat{M}_i - M_i)^2}
\]  

(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 \textit{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 \( \times \) 8 m \( \times \) 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; \( D_{max} \) is taken 3 m. The camera’s pose is sampled at 12° intervals. The cameras have the same horizontal and vertical viewable angles, \( \phi_h, \phi_v \), of 60° and have a working distance, \( r \), of 7 m. The maximum visibility angle, \( \theta_{max} \), (2) and the angle, \( \phi_{max} \), (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

<table>
<thead>
<tr>
<th>IN</th>
<th>$AF_{con}$</th>
<th>$B_a$</th>
<th>$B_b$</th>
<th>$\Delta_{rec}$</th>
<th>$\Delta Z_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>100</td>
<td>80</td>
<td>3.3</td>
<td>4.5</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>70</td>
<td>60</td>
<td>4.5</td>
<td>6.2</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>60</td>
<td>50</td>
<td>4.9</td>
<td>7.9</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>50</td>
<td>40</td>
<td>5.2</td>
<td>9.7</td>
</tr>
</tbody>
</table>

IN: the index number; $AF_{con}$: the reconstruction accuracy requirement; $B_a$, $B_b$: baseline lengths for each pair, [cm]; $\Delta_{rec}$: the maximum value reconstruction error of pairs, [cm]; $\Delta Z_{max}$: 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, $\psi$, 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


An Adaptive Quality Assessment System
– Aspect of Human Factor and Measurement Uncertainty

Authors:
Jenny Wirandi, Jiandan Chen and Wlodek Kulesza

Reformatted version of paper accepted to:
IEEE Transactions on Instrumentation and Measurement, January, 2008 (Printing)
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 cannot 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, \([\alpha]\), 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 \textit{quantitative features} and the \textit{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 \textit{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 \textit{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 $[\alpha]$.
2. \textit{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.](image)

The weights are calculated from quantitative factors. For instance, they could be calculated as a mean value or product of the measurement uncertainty ratio, $\alpha_{ui}$ and relevance ratio, $\alpha_{ri}$, e.g.:

$$\alpha_i = \frac{\alpha_{ui} + \alpha_{ri}}{2}$$

(1)

where the measurement uncertainty ratio $\alpha_{ui}$ is calculated as:

$$\alpha_{ui} = 1 - u_c(y_i)$$

(2)

where $u_c(y_i)$ is the combined relative standard uncertainty of output estimates, $y_i$, [3].

The relevance ratio, $\alpha_{ri}$, 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, $H$. 

The QIs have been estimated by a simplified image quality system.
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.

\[ \alpha_{ri} = \frac{h_i}{H} \]  

\[ (3) \]

**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:
where $x$ denotes the intensity data of the image being tested, and $y$ denotes the intensity data of the image pattern; $\sigma_x$, $\sigma_y$ are the standard deviations related to the contrast of the test and pattern images respectively; $\sigma_{xy}$ 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, $\alpha$:s, are based on the measurements’ uncertainty and features’ relevance, [15]:

$$F_{IQ} = (F_x)^\alpha \cdot (F_y)^\alpha \cdot (F_c)^\alpha$$

For reasons of simplification, in this case study all weights, $\alpha$: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’ 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
The NN was implemented with the help of the Matlab Neural Network Toolbox and the three-layer \texttt{transig/transig/logsig} network with ten neurons in each layer [16]. Afterwards, the Back-Propagation approach Neural Network was applied. The training function \texttt{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 \textit{et al.} [14], [15], it is clear that our QIs predict the human assessment with much better accuracy, (see Fig. 8).
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.
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.
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 $l_{\text{target}}$, $t_{\text{target}}$ and $b_{\text{target}}$), 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.
where $\sigma_l$, $\sigma_t$, $\sigma_b$ are the standard deviations for test fibre length, tear index and brightness respectively; $\sigma_{\text{target}}$, $\sigma_{\text{target}}$, $\sigma_{\text{target}}$ are the standard target deviations for target fibre length, tear index and brightness respectively; $l$, $t$ and $b$, are the mean values of fibre length, tear index and brightness respectively, and $l_{\text{target}}$, $t_{\text{target}}$, $b_{\text{target}}$ 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, $|F_i^{(j)}| \leq 1$. As with the image quality model, the product of all exponentially weighted features, $F_{PQ}$, is used:

$$F_{PQ} = \left( F_{l}^{(1)} \right)^{\alpha_{l}^{(1)}}, \left( F_{l}^{(2)} \right)^{\alpha_{l}^{(2)}}, \left( F_{t}^{(1)} \right)^{\alpha_{t}^{(1)}}, \left( F_{t}^{(2)} \right)^{\alpha_{t}^{(2)}}, \left( F_{b}^{(1)} \right)^{\alpha_{b}^{(1)}}, \left( F_{b}^{(2)} \right)^{\alpha_{b}^{(2)}}$$

where $|\alpha_{j}^{(0)}| \leq 1$.

The weights, $\alpha_{j}^{(0)}$'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, $a_j^B$, 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


ABSTRACT

In our aging society, the design and implementation of a high-performance autonomous distributed vision information system for autonomous physical services become ever more important. In line with this development, the proposed Intelligent Vision Agent System, IVAS, is able to automatically detect and identify a target for a specific task by surveying a human activities space. The main subject of this thesis is the optimal configuration of a sensor system meant to capture the target objects and their environment within certain required specifications. The thesis thus discusses how a discrete sensor causes a depth spatial quantisation uncertainty, which significantly contributes to the 3D depth reconstruction accuracy. For a sensor stereo pair, the quantisation uncertainty is represented by the intervals between the iso-disparity surfaces. A mathematical geometry model is then proposed to analyse the iso-disparity surfaces and optimise the sensors’ configurations according to the required constraints. The thesis also introduces the dithering algorithm which significantly reduces the depth reconstruction uncertainty. This algorithm assures high depth reconstruction accuracy from a few images captured by low-resolution sensors. To ensure the visibility needed for surveillance, tracking, and 3D reconstruction, the thesis introduces constraints of the target space, the stereo pair characteristics, and the depth reconstruction accuracy. The target space, the space in which human activity takes place, is modelled as a tetrahedron, and a field of view in spherical coordinates is proposed. The minimum number of stereo pairs necessary to cover the entire target space and the arrangement of the stereo pairs’ movement is optimised through integer linear programming.

In order to better understand human behaviour and perception, the proposed adaptive measurement method makes use of a fuzzily defined variable, FDV. The FDV approach enables an estimation of a quality index based on qualitative and quantitative factors. The suggested method uses a neural network as a tool that contains a learning function that allows the integration of the human factor into a quantitative quality index. The thesis consists of two parts, where Part I gives a brief overview of the applied theory and research methods used, and Part II contains the five papers included in the thesis.

Keywords: 3D Reconstruction, Iso-disparity Surfaces, Depth Reconstruction Uncertainty, Uncertainty Analysis, Dither, Sensor Placement, Multi Stereo View, Image Quality, Human Factor.