10

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

Wlodek Kulesza, Jiandan Chen and Siamak Khatibi
Blekinge Institute of Technology
Sweden

1. Introduction

The requirements for autonomous physical services supporting people have become more important in recent years in activities such as taking care of the elderly people, doing the housework, and giving a comfortable living environment. For this reason, our research is aimed towards design and implementation of a high-performance autonomous, distributed vision information system, which would be able to understand human behaviours and living environment, as a temporary substitute for a qualified nurse and housekeeper.

The human-centred computation is proposed in the MIT Oxygen Project, (MIT, 2008). Furthermore, Hashimoto presented 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*, (Hashimoto, 2003). The Intelligent Space was treated as a platform, which supported people's information and physical needs. It was the interface for both the humans and robots.

The proposed Intelligent Vision Agent System, IVAS, is a high-performance, autonomous, distributed vision and information processing system. Figure 1 illustrates the idea of the IVAS. It consists of multiple sensors and actuators for surveillance of the human activities space involving humans and their surrounding environment including robots and household appliances etc. The system not only gathers information, but also controls these sensors including their deployment and autonomous servo. But first of all it is able to extract required information from images for different applications, especially for three dimensional (3D) reconstruction. 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 cameras to get optimised 3D information. The intelligent agent consists of a knowledge database, with learning and decision making components that can be used to track, recognize and analyze the objects.

Similar to the human eyes, the 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. 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, (Völpel & Theimer, 1995). The shape of iso-disparity surfaces for general stereo configurations was studied by Pollefeys & Sinha (Pollefeys &

Source: Stereo Vision, Book edited by: Dr. Asim Bhatti,
ISBN 978-953-7619-22-0, pp. 372, November 2008, I-Tech, Vienna, Austria

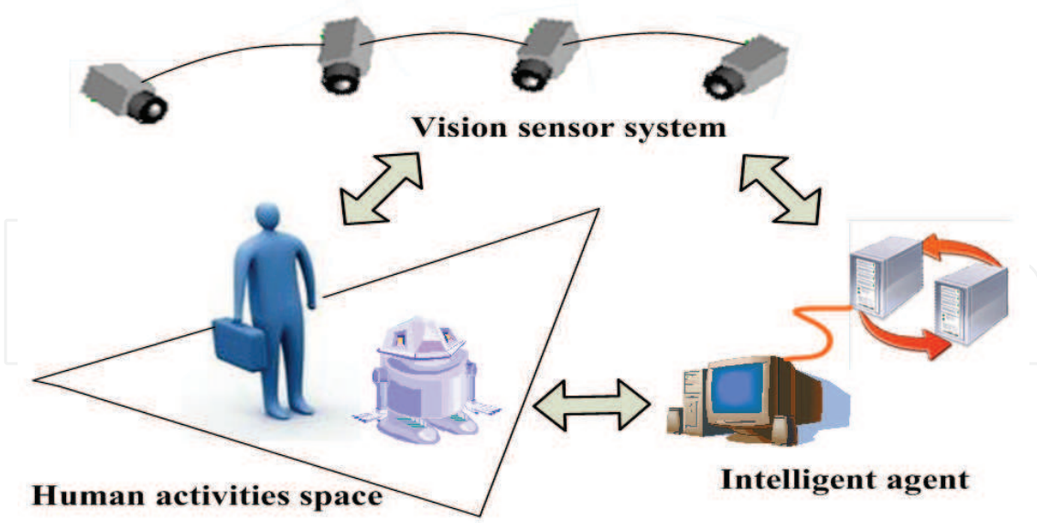


Fig. 1. Overview of an Intelligent Vision Agent System

Sinha , 2004) and Chen et al. (Chen et al., 2007c). The proposed mathematical model of iso-disparity map provides an efficient way of describing the shape of iso-disparity surfaces, and estimating the depth reconstruction uncertainty which is related to the stereo pair baseline length, the target distance to baseline, focal length, convergence angle and the pixel resolution.

The depth spatial quantization uncertainty, caused by a discrete sensor is one of the factors which has the most influence on the depth reconstruction accuracy. This type of uncertainty cannot be decreased by the reduction of pixel size since the signal to noise ratio is also reduced and the sensitivity of the sensor itself is restricted. The selection of an optimal sensor pixel is discussed by Chen et al., (Chen et al., 2000).

The sensor planning by re-annealing software was introduced by Mittal, (Mittal, 2006) and the evaluation of the sensors' configuration by a quality metric was presented in (Chen, 2002). A linear programming method to optimize sensor placement based on binary optimization techniques has been developed, (Chakrabarty et al., 2002; Hörster & Lienhart, 2006; Erdem & Sclaroff, 2006). This is a convenient tool for optimising the visual sensors' configurations when observing target space such as human activities space. Chen at al. described the optimization program for the 3D reconstruction of a human activity space, (Chen et al., 2007a; Chen et al., 2007b). The papers introduce the method of optimizing stereo pair configurations under the required constraints of stereo pair baseline length, visibility, camera movement and depth reconstruction accuracy.

The first part of this chapter introduces a mathematical geometry model which is used to analyze 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 3D space in order to get sufficient visibility and accuracy for surveillance, tracking and 3D reconstruction. The depth reconstruction accuracy is quantitatively analyzed by the proposed model. The proposed iso-disparity mathematical model presents possibility of reliable control of the iso-disparity curves' shapes and intervals by applying the systems configuration and target properties.

In the second part of this chapter, the key factors affecting the accuracy of 3D reconstruction are analysed. It shows that the convergence angle and target distance influence the depth

reconstruction accuracy most significantly. The depth accuracy constraints are implemented in the model to control the stereo pair's baseline length, position and pose. It guarantees a certain accuracy in the 3D reconstruction. The reconstruction accuracy is verified by a cubic reconstruction method. The optimization is implemented by applying the camera, object and stereo pair constraints into the integer linear programming.

2. Quantized depth reconstruction uncertainty

Two images are needed which are 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. There are two terms related to the depth reconstruction: disparity and quantized depth reconstruction uncertainty. *Disparity* in our approach refers to the displacement of corresponding points along the corresponding epipolar lines for a common scene point, (Pollefeys & Sinha, 2004). In the case where epipolar lines are horizontal, the disparity is measured directly from the difference between the co-ordinates of the corresponding points. The inverse projection of all possible image points with the same disparity will allow reconstruction of the iso-disparity surfaces in 3D space. *Quantized depth reconstruction uncertainty* is defined as the interval between discrete iso-disparity surfaces due to the discrete sensor.

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

There are two configurations for a stereo pair in common use. The first simple configuration 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 the cameras' field of view, FoV. The second common configuration is the convergence stereo pair, where the optical axes cross at a fixation point. The simple mathematical models of the iso-disparity map for these configurations are analyzed in the following subchapters.

2.1 The iso-disparity map of a parallel stereo pair

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

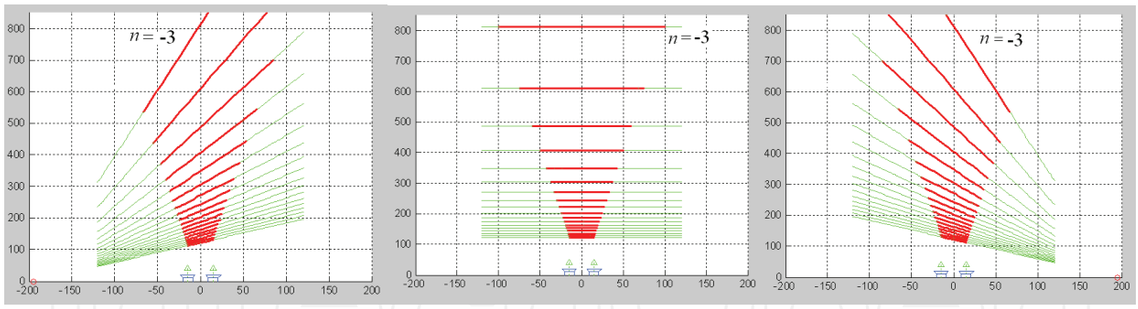
$$z(x,n) = \frac{f_L - f_R}{n\Delta D}x + \frac{B}{2n\Delta D}(f_L + f_R) \tag{1}$$

where B is the baseline length, n is an integer and ΔD is the disparity resolution.

The planes are shown as the thin green lines in Figure 2(a) and Figure 2(c). All the iso-disparity planes intersect with the xy-plane (the stereo pair baseline is a part of the x-axis), and converge on a straight line:

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

It is clear from equation (1) that when the focal lengths are equal $f_L=f_R=f$, z becomes independent of x and the iso-disparity planes are parallel to the xy-plane, see the green lines in Figure 2(b).



(a) (b) (c)

Fig. 2. Iso-disparity planes for parallel stereo pairs from the synthetic simulation (the red lines) and a plot of the mathematical model from equation (1) (the green lines). The lines are plotted with steps of 10 pixels. (a) Cameras with different focal lengths, f_L=3.5 cm, f_R=3.0 cm for left and right camera respectively. The convergence point is (-195 cm, 0) on the xz-plane. (b) Cameras with the same focal length of 3.25 cm. (c) Cameras with different focal lengths f_L=3.0 cm, and f_R=3.5 cm for left and right cameras respectively, the convergence point is (195 cm, 0) on the xz-plane.

From the inverse projection of the image points and by applying the triangulation method, using the Epipolar Geometry Toolbox, (Mariottini & Prattichizzo, 2005), we can get the synthetic iso-disparity surfaces. Figure 2 shows the synthetic disparity surfaces (the red lines) and the plots from (1) (the green lines). Here the baseline length B is 30 cm and the disparity resolution ΔD=0.04 cm, or ten sensor pixel lengths where p=0.004 cm. Figure 2(a) and Figure 2(c) are plotted for the parallel stereo pair with different focal lengths. The parallel iso-disparity planes for parallel stereo pairs with the same focal lengths are shown in Figure 2(b). The synthetic simulations and the results calculated from (1) give similar results.

2.2 The iso-disparity surface of a convergent stereo pair

Let us consider two cameras with a convergence angle α_c, where $\alpha_{cL0}=\alpha_{cR0}=\alpha_c$ for the left and right camera respectively, with the angles rotated inwards to achieve a fixation point FP_0, as in Figure 3. If the point TP_0 lies on the baseline's axis of symmetry, then the angles, (ψ_{L0}, ψ_{R0}), are the angles between the visual lines and a line perpendicular to the baseline. The zero disparity circle is defined by the fixation point and the left and right camera position points C_L and C_R. This circle is known as the Vieth-Müller circle, and is a projection of the horopter, (Ogle, 1950).

The iso-disparity surface is a cylinder whose cross section on the xz-plane is a conic section, which passes through the centers of projection C_L and C_R, and the point M_∞. The point M_∞ is a point imagined at infinity in both images, which can be obtained from the intersection of the normals to the optical axes, going through the projection centres, (Pollefeys & Sinha , 2004). It is possible to prove that for the case when $\alpha_{cL0}=\alpha_{cR0}=\alpha_c$, the conic is an ellipse. To determine the ellipse we need to estimate its five degrees of freedom. Three of these are determined by the points C_L, C_R and M_∞. One of the two remaining degrees is related to the point TP_0 with the disparity $n\Delta D$. The relationship between disparity $n\Delta D$ and focal lengths f_L and f_R for the left and right cameras respectively, is the last required degree of freedom. If the disparity $n\Delta D$ and focal lengths f_L and f_R are known, the unique ellipse can be determined.

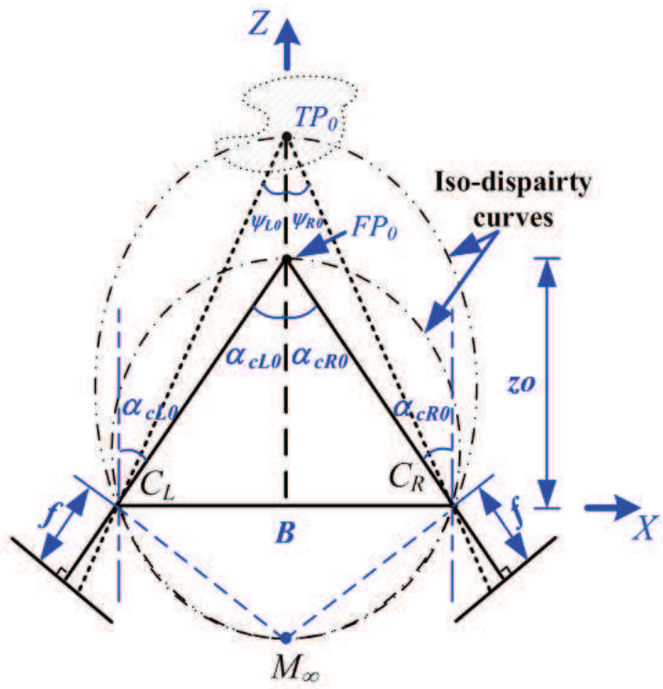


Fig. 3. An example of the iso-disparity curves for the convergence stereo pair in the plane defined by the cameras optical axes. z_0 is the distance from the fixation point to the baseline, f is the focal length.

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

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

For the chosen co-ordinates $x_{0e}=0$ and $z_{0e}=b-B\tan\alpha_c/2$, then:

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

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

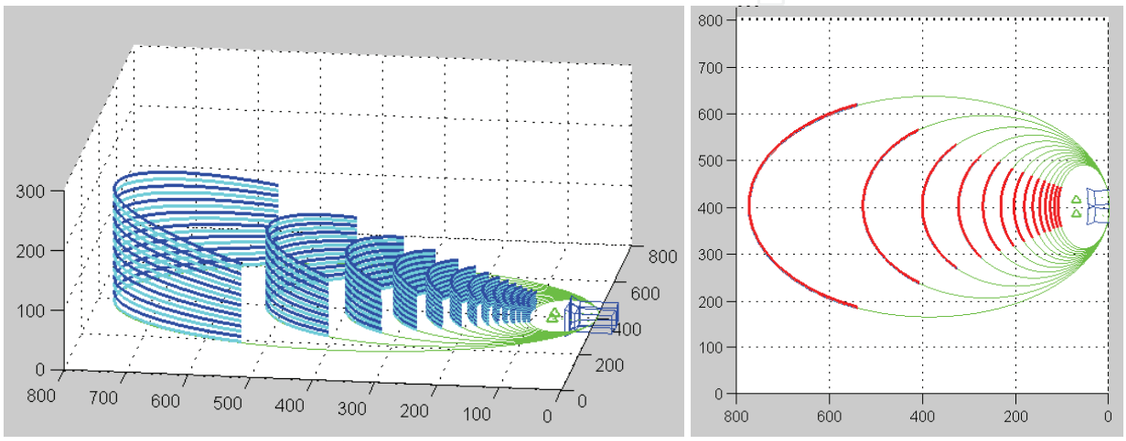
$$b=\frac{\frac{B}{2}}{\sin 2\alpha_c-\frac{n\Delta D}{f}\cos^2\alpha_c} \tag{5}$$

The ellipse half-axis along the x-axis, a, can be found from the relationship:

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

where $\psi_c = \psi_{L0} = \psi_{R0}$.

Figure 4 shows the iso-disparity surfaces for a stereo pair from the synthetic simulations (bold blue and red lines) and from theoretical model (4) (thin green and light blue lines) in 3D space, in perspective view, Figure 4 (a), and top view, Figure 4 (b). The characteristics of cameras are: a baseline length B=50 cm, convergence angle, α_c=4°, focal length f=2.5 cm and disparity resolution ΔD=0.04 cm. Both results match each other perfectly.



(a) (b)

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

2.3 Mapping of 2D uncertainty for a stereo pair configuration

Since the gaps between iso-disparity surfaces represent the quantization 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 to optimise the configuration of the stereo setup. Mapping of the 2D uncertainty 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, (Chen et al., 2007a). 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 then five points are needed. Two of these points can be the grid point and its symmetrical point, with respect to the symmetry axis of the baseline. The three others points are C_L, C_R and M_∞. For a convergent stereo pair, the ellipse axes a and b can be found using the ellipse fitting algorithm, (Halif & Flusser, 1998). Then using equation (5), the two closest

ellipses with discrete disparity values $n\Delta D$ and $(n+1)\Delta D$ respectively, can be found, where the disparity resolution ΔD is one sensor pixel length.

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

2.3.1 Simulation results

The presented simulations 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 z for a given stereo baseline length B, focal length f, and sensor pixel length p. The results are presented in Figure 5, where the cameras optical axes are in the xz-plane. The depth reconstruction uncertainty is specified by the positive y-axis of the co-ordinate. 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

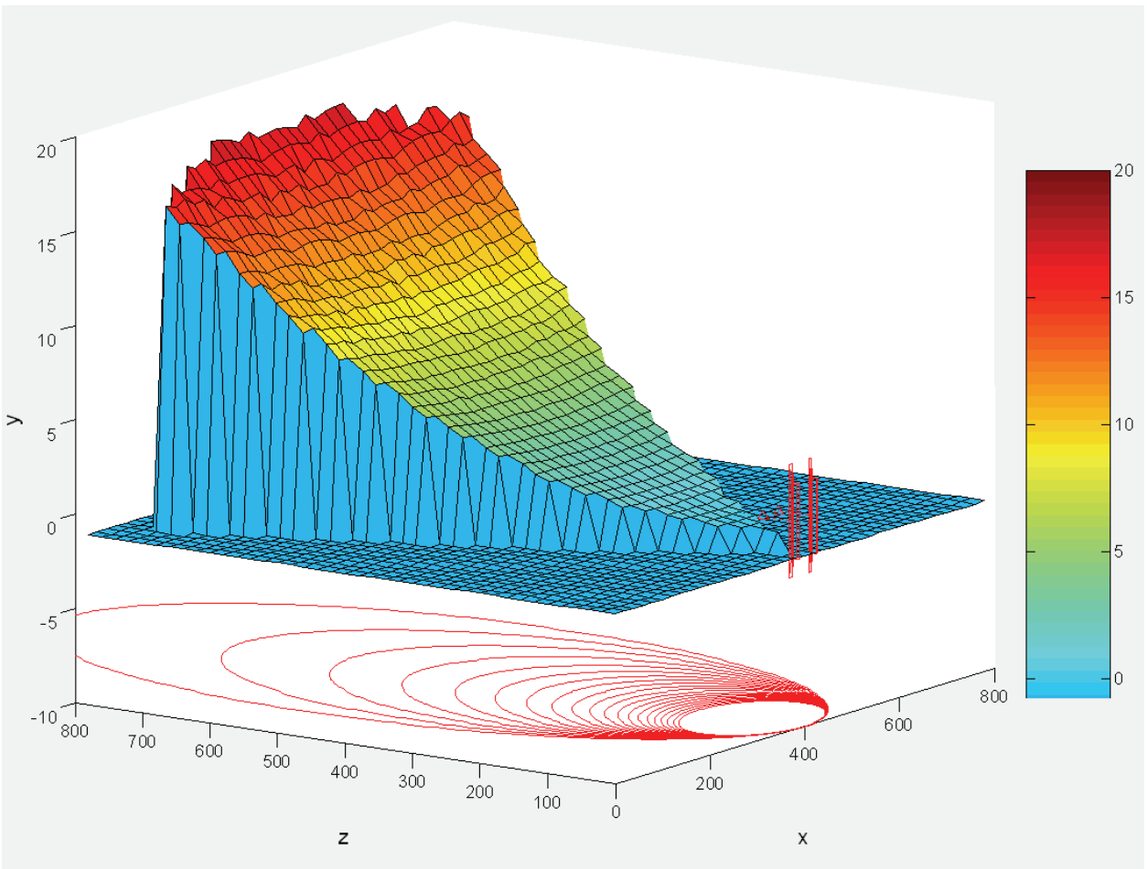


Fig. 5. The depth reconstruction uncertainty map for a stereo pair's FoV, where B=40 cm, f=3.5 cm and p=0.004 cm.

generated with baseline length B=40 cm, focal length f=3.5 cm and pixel length p=0.004 cm, stereo convergence angle, α_c=4° and the FoV is approximately 54°. This case study proves that the quantized depth reconstruction uncertainty increases as the distance to the target increases. To show the quantized properties of depth reconstruction uncertainty, the map of the iso-disparity curves with suitable pixel length is shown in Figure 6. The figure shows only half of the FoV, with a cross section along the ellipses' axes perpendicular to the baseline. The quantization step increases with the target distance.

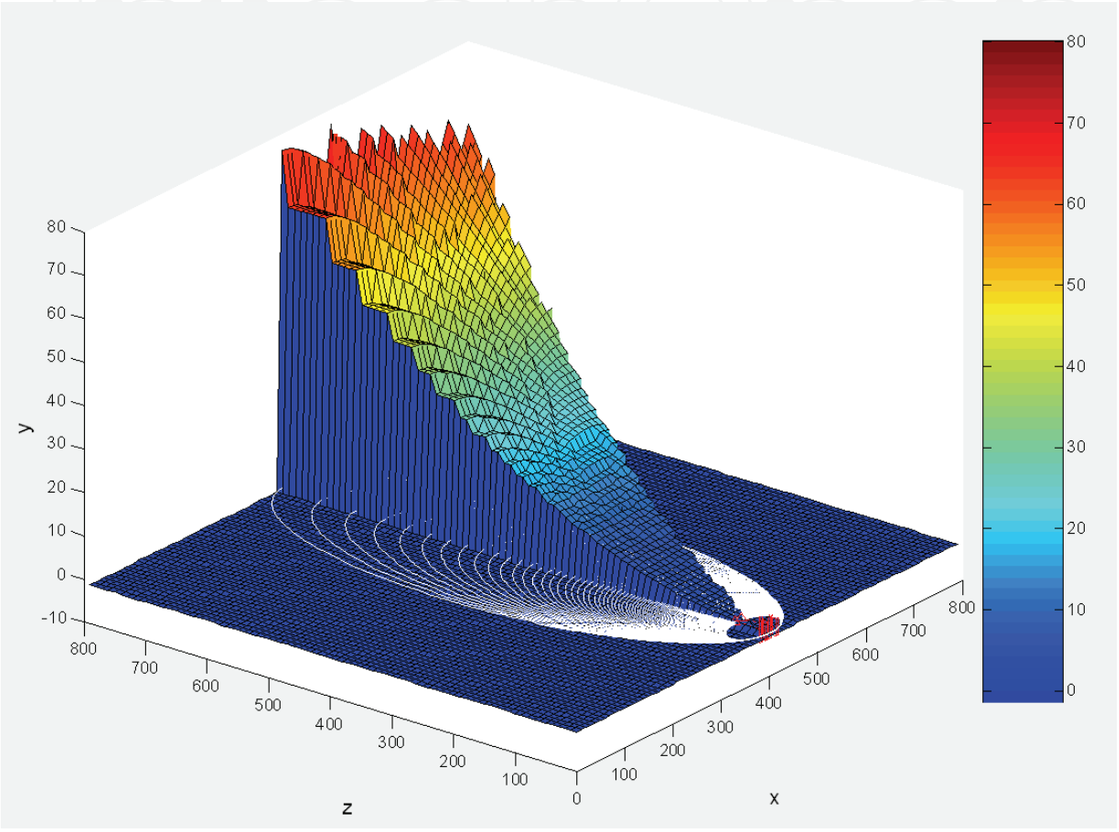
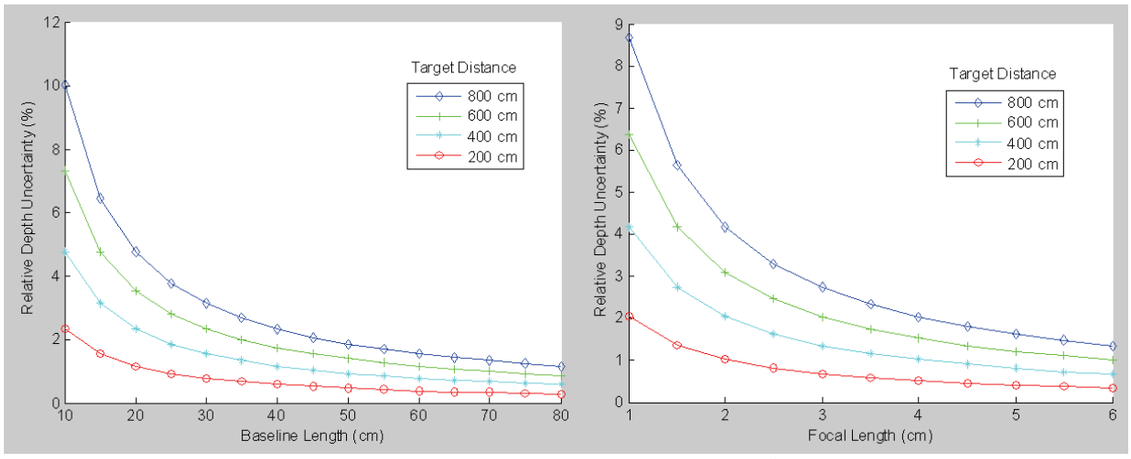


Fig. 6. The depth reconstruction uncertainty map for a stereo pair's half FoV, where B=20 cm, f=3.5 cm and p=0.008 cm.

Exact illustrations of how the depth reconstruction uncertainty varies with the baseline lengths, focal lengths, sensor pixel length and stereo convergence angle, are shown in Figure 7 and Figure 8. Figure 7(a) shows that the relative depth reconstruction uncertainty, related 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.

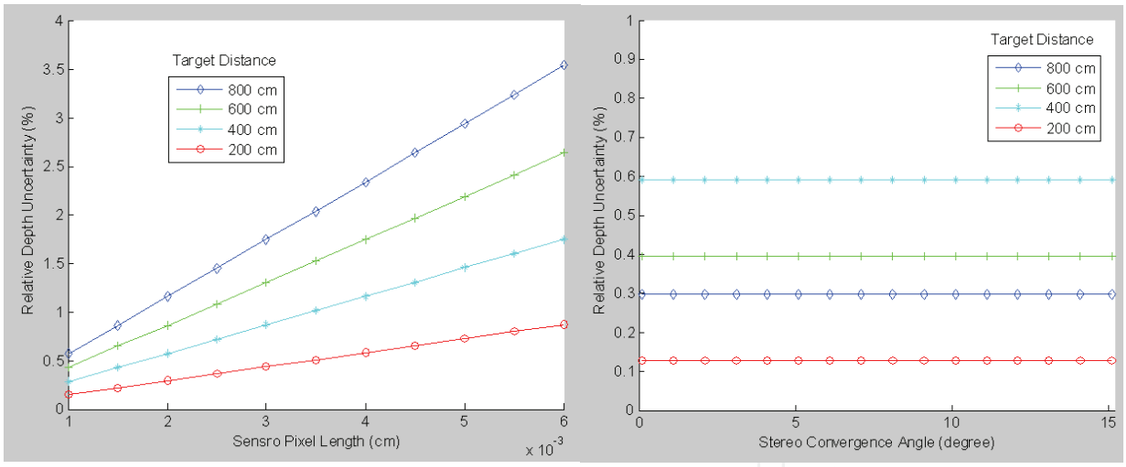
The change of the relative depth reconstruction uncertainty versus the focal length is similar to that of the baseline length; see Figure 7(b). For a focal length of longer than 3.5 cm the increase of the uncertainty is relatively slow. Its minimum tends to be consistently between 1.5% and 0.4% for target 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.



(a) (b)

Fig. 7. The relative quantized depth uncertainty as function of the baseline length (a), focal length (b). The distance from the target to the camera is 800 cm, 600 cm, 400 cm and 200 cm, respectively.

Furthermore, Figure 8(a) illustrates the linear 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 also depends on the target distance. Figure 8(b) shows that the stereo convergence angle has a slight influence on the uncertainty but this also depends on the target distance.



(a) (b)

Fig. 8. The relative quantized depth uncertainty as a function of the sensor pixel length (a), and stereo convergence angle (b). The distance from the target to the camera is 800 cm, 600 cm, 400 cm and 200 cm, respectively.

Figure 9(a) and Figure 9(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. Also, a significant increase in the uncertainty is visible for a focal length below 3.5 cm.

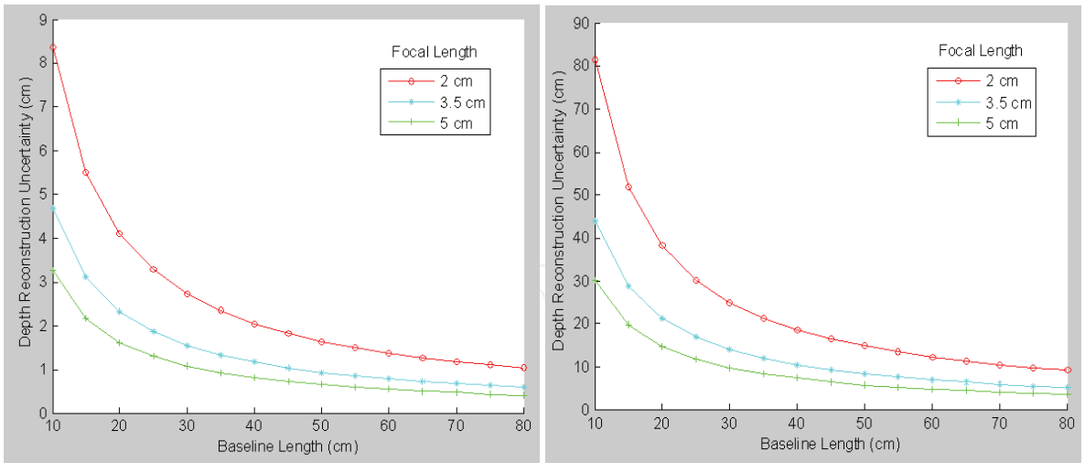


Fig. 9. The relative quantized depth uncertainty varies with both the focal length and the baseline length. The focal lengths are 2 cm, 3.5 cm and 5 cm, respectively, marked by different type of lines. The target is (a) 200 cm; (b) 600 cm far away from the camera.

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

3. Arrangement of a multi stereo visual sensor system

Human activity 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 arrangement can be viewed as an extension to the well-known Art Galley Problem, AGP, (O'Rourke, 1987). 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. In our approach, a similar task is required to find the minimum number of stereo pair sensors needed to cover the target space. Here the human activities space as a target space is defined by a tetrahedron.

This subchapter gives an overview of the theory for the multi stereo sensors arrangement in the intelligent vision system. In the (Chen et al., 2007a; Chen et al., 2007b), we suggested the camera constraints, which focused on the visibility to the target. The accuracy constraint is based on the estimation of quantized depth reconstruction accuracy, where a target object has the same target convergence angle. The iso-disparity geometry model gives a deep analysis of depth reconstruction accuracy. It analyzes the whole camera FoV. This can be used to dynamically adjust the position, poses and baselines length of multiple stereo pairs of cameras in order to get the desired accuracy.

The planning algorithm proposed in the (Chen et al., 2007a; Chen et al., 2007b) works in 3D space. The approach dynamically adjusts the stereo pair baseline length according to the accuracy requirement and the target distance as a distance from the target position to stereo pair baseline. The minimal amount of stereo pairs to cover human activity space is solved by means of Integer Linear Programming, ILP, (Chakrabarty et al., 2002; Hörster & Lienhart,

2006; Berkelaar et al., 2005). The 3D reconstruction accuracy, which is ensured by an accuracy constraint, can be verified by a cubic reconstruction.
The constraints of the stereo view optimization model can be defined from the environment, camera properties and human behaviour, which affect the process of identification and reconstruction of the target.

3.1 Constraints for the optimization model

The human activities space is modelled by a tetrahedron as shown in Figure 10. The normal of each tetrahedron's upper triangle gives the orientation of that surface. If the visibility angle, θ, between the triangle normal and a line drawn from the centroid of the triangle to the camera position, increases then the image resolution decreases. In order to get a good image resolution, a visibility angle, θ, of less than the maximum visibility angle, θ_{max}, is required which can be expressed as constrain:

$$\theta \le \theta_{\max} \tag{7}$$

The camera orientation should line up with the centroid of the triangle, thus bringing the target object to the centre of 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 the triangle, φ, of less than the maximum angle φ_{max} is required:

$$\varphi \le \varphi_{\max} \tag{8}$$

In order to follow the movement of the target object, a camera 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, formulated as the camera maximum movement should be less than the maximum distance of camera movement which the system supported.

$$Dist(StereoPair_{next}, StereoPair_{current}) \le Dis_{\max} \tag{9}$$

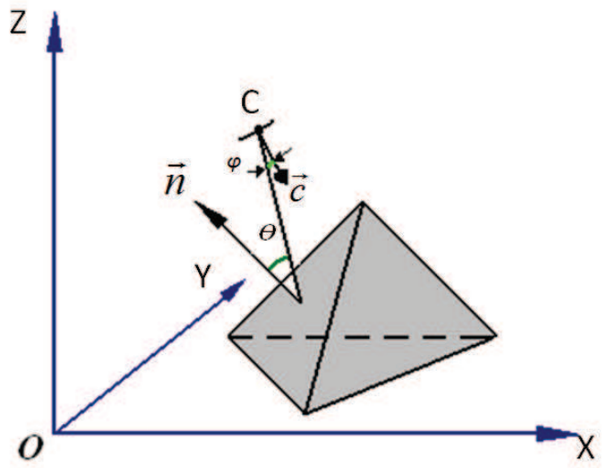


Fig. 10. Illustration of the human space modelled as a tetrahedron; θ - the visibility angle between the triangle normal $\vec{c}$ and a line from the centroid of the triangle to the camera position; φ - the angle between the camera orientation $\vec{c}$ and a line from the camera position to the centroid of the triangle.

A fewer number of potential next-view positions for the cameras restricted by (9)) can simplify computation.

The camera constraints are related to the camera FoV. The camera horizontal and vertical viewable angles, ϕ_h, ϕ_v, and a working distance, r, can be calculated from the camera attributes, see the spherical co-ordinate systems shown in Figure 11. In order to cover the target object feature points by the camera FoV, the following constraints must be fulfilled:

$$\begin{aligned} & l \le r \text{ and} \\ & \alpha_c - \phi_h/2 \le \alpha_o \le \alpha_c + \phi_h/2 \\ & \beta_c - \phi_v/2 \le \beta_o \le \beta_c + \phi_v/2 \end{aligned} \tag{10}$$

where: l is the distance between the target position and camera's position; α_o, β_o are respectively the azimuth and elevation of target; α_c, β_c are respectively the azimuth and elevation of the camera's pose.

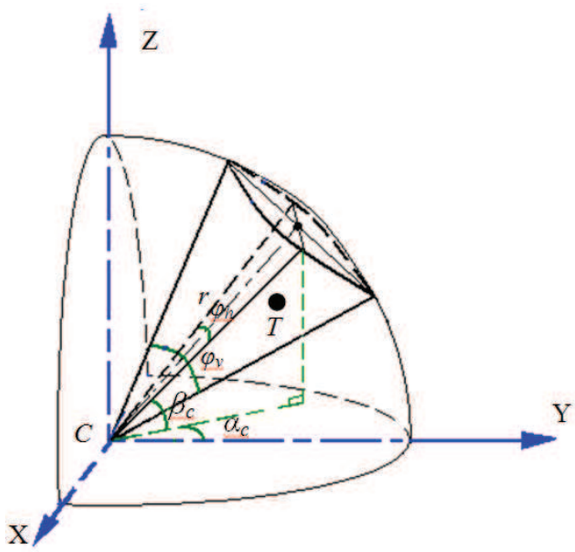


Fig. 11. The spherical co-ordinates system and FoV of a camera where C is camera position and the example target point located at point T.

Since stereo matching becomes more difficult when the baseline distance increases, the baseline length B has to be limited to the maximum stereo baseline length, B_{max}:

$$0 < B \le B_{\max} \tag{11}$$

The depth reconstruction is one of major focus in this research project. The depth reconstruction accuracy improvement can be adjusted by the baseline length, (Mittal, 2006; Samson et al., 2006). This research introduces the concept of a depth accuracy factor, AF, which is a function of target convergence angle, ψ, and camera pose, α_c. 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 of the target objects. In the most common case, the cameras form a converging stereo pair. Cameras poses azimuths and baseline are shown in Figure 12. Cameras convergence angles, (α_{cl}, α_{cr}), are the angles of each camera rotated inwards from the parallel to achieve

convergence. The target convergence angles, (ψ_l, ψ_r), are the angles between the visual lines of each camera and the baseline perpendicular. From Figure 12, simplifying: $\psi = \psi_l = \psi_r$ and $\alpha_c = \alpha_{cl} = \alpha_{cr}$, we obtain:

$$Z = \frac{B}{2\tan\psi} \tag{12}$$

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

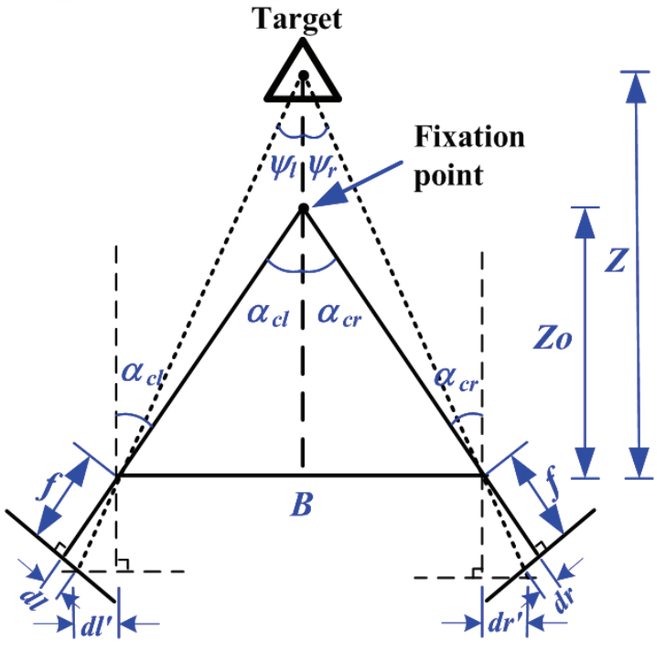


Fig. 12. Model of stereo pair geometry; Z_0 is the depth of the fixation point, $|dl\text{-}dr|$ is stereo disparity, f is focal length.

The equation (12) can be written as:

$$Z = \frac{B}{2}\cdot\left(\frac{1}{\tan\alpha_c - \tan(\alpha_c - \psi)} + \frac{\tan\alpha_c \cdot \tan(\alpha_c - \psi)}{\tan\alpha_c - \tan(\alpha_c - \psi)}\right) \tag{13}$$

In the case of parallel stereo or with the target close to the fixation point, the α_c or (α_c-ψ) varies by a small amount, and the equation (13) can be further simplified. The resolution of the target convergence angle, ψ, is related to a single pixel, p, in the image, thus the relative depth error can be written as:

$$\frac{\Delta Z}{Z} \approx \left|\frac{\cos\alpha_c}{\sin\psi\cos(\alpha_c - \psi)}\right| \cdot \frac{p}{f} = AF \cdot \frac{p}{f} \tag{14}$$

where AF is the depth accuracy factor, f is a focal length, and the depth quantization error is assumed to be one pixel, p.

The depth error is proportional to the depth accuracy factor. In fact, since the depth accuracy factor varies more significantly with respect to the target convergence angle, ψ,

than to the camera's pose, α_c, the target convergence angle determines the depth accuracy factor. The accuracy constraint for a given p can be defined as:

$$AF \le AF_{con} \tag{15}$$

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

3.2 Implementation of the camera arrangement

The stereo pair arrangement consists of three stages:

- Firstly, we find potential stereo pairs that satisfy stereo constraints by greedy algorithm from all potential cameras' positions and poses.
- Secondly, the integer linear programming is applied to minimize the total amount of stereo pairs subject to the visibility and baseline length constraints, depth accuracy constraints and camera movement constraints. The 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.
- Finally, the 3D reconstruction accuracy can be verified by a cubic reconstruction simulated using a pair of rectified scene images.

3.2.1 Greedy algorithm

This algorithm gives a flexible way of organising 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 positions $C_n(x_{cn},y_{cn},z_{cn})$ and poses $\psi_n(\alpha_{cn},\beta_{cn})$ of the camera state, $Scamera^{k}_{C_n,\psi_n}$, where k is the camera state index number. The target object, which must be covered, is modelled as a tetrahedron. In the next step, we calculate all of the potential camera positions and poses needed to cover each upward triangle of this model. Taking this, we combine every two camera states to be a potential stereo pair, $Stereopair_i$, according to the stereo constraint (Chen et al., 2007a; Chen et al., 2007b). The algorithm is sufficiently flexible to add other constraints for stereo pair, e.g. the angle constraint between the cameras' optical axes. Finally the algorithm removes the redundant potential stereo pairs.

3.2.2 Integer linear programming

This model assumes that one type of camera is used throughout, resulting in just one camera FoV being considered. The optimization of the amount of cameras with different FoVs can also 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 calculated and stored in advance. The stereo visibility binary variable table $Stereovis_{j,i}$ is defined by:

$$Stereovis_{j,i} = \begin{cases} 1 & \text{if a } Stereopair_i \text{ covers} \\ & \text{triangle } j \text{ of target object model} \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

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

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

$$\min \sum_{i=1}^{K_s} S_i \tag{17}$$

subject to

$$\sum_{i=1}^{K_s} S_i \times Stereovis_{j,i} \ge 1, \quad \text{for } j = 1,2,3 \tag{18}$$

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:

$$c_{p,i} = \begin{cases} 1 \text{ if two pairs } i \text{ and } p \text{ share the same camera} \\ \quad \text{with different orientations, where } i \ne p \\ 0 \text{ otherwise} \end{cases} \tag{19}$$

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

One more constraint is added into the model:

$$\sum_{i=1}^{K_s} S_i \times c_{p,i} \le 1, \quad \text{for } p = 1, 2, \cdots, K_s \tag{20}$$

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

3.2.3 Cubic reconstruction

The rectification matrix is calculated directly from the **p**erspective **p**rojection **m**atrix, PPM, (Fusiello et al., 2000), and the rectification algorithm also gives two new PPMs, P_n1 and P_n2. The cubic reconstruction in 3D can be performed with a triangulation method directly from the rectified images, using P_n1, P_n2. The 3D reconstruction error, Δ_{rec}, for a single pixel error along a horizontal direction in the rectified image, has same value as the depth error and is given by:

$$\Delta rec = \sqrt{\frac{1}{8}\sum_{i=1}^{8}\left|\hat{M}_i - M_i\right|^2} \tag{21}$$

where $\hat{M}_i$ gives the co-ordinates 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.

3.3 The model validation

The simulations were performed in MATLAB 7.0. The integer linear programs *lpsove package,* (Berkelaar et al., 2005), and the Epipolar Geometry Toolbox, (Mariottini & Prattichizzo, 2005), were used to minimise the number of cameras and transform the object position in 3D. The simulation environment consists of a rectangular room of 8 m × 8 m × 3 m. The modelling of the human activities space as a tetrahedron requires three upward triangles; and each triangle must be visible to at least one pair of cameras. Each model is 2 m high and 1.2 m at the base edges. The cameras' positions are restricted to the ceiling around the room, their potential positions sampled at 0.2 m intervals in the initial phase, and 0.1 m intervals for the next cameras viewing position; Dis_{max} is taken 3 m. The camera's pose is sampled at 12° intervals. The cameras have the same horizontal and vertical viewing angles, ϕ_h, ϕ_v, of 60° and have a working distance, r, of 7 m. The maximum visibility angle, θ_{max}, (7)) and the angle, φ_{max}, (8) are taken to be 70° and 10° respectively. The pixel size of our vision system, p, is 0.02 mm and the 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 the 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 Figure 13. The arrows indicate the optical axes of the cameras.

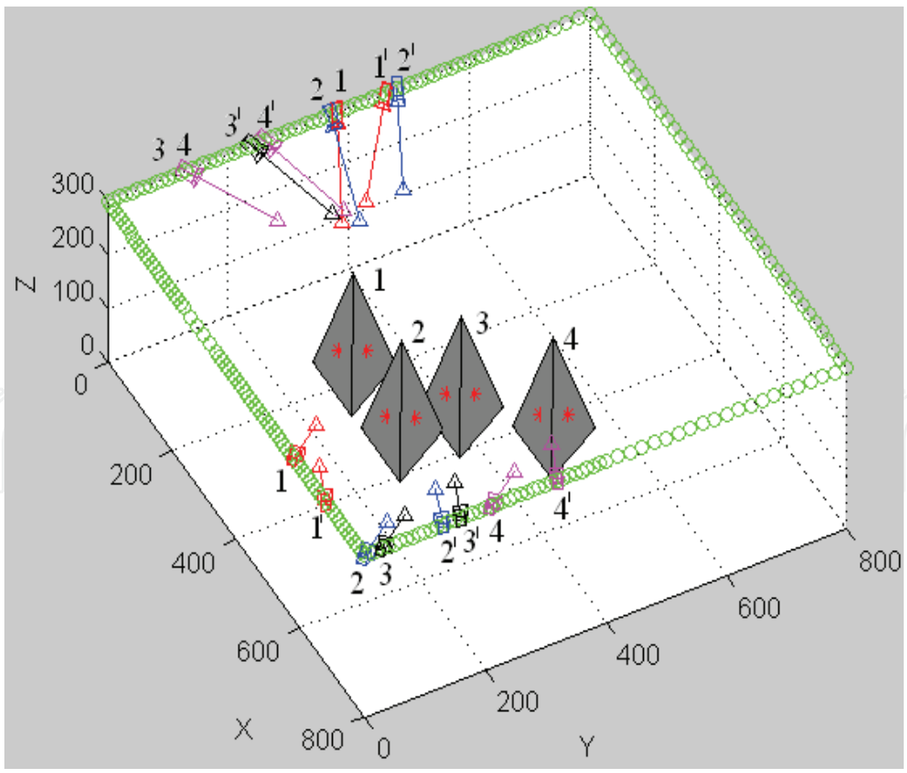


Fig. 13. The stereo pair positions, poses and baselines with the depth accuracy factor AF_{con} = 8, for the moving target.

The index numbers indicate the model locations and corresponding camera 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 previous position with the constraint (9). In each position every upward triangular view 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 changes dynamically according to the distance to the target.

Figure 14 illustrates a case of four different values of AF_{con} applied 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 Tabble 1. It proves that the baseline increases as AF_{con} becomes more restricted, and the reconstruction error is smaller.

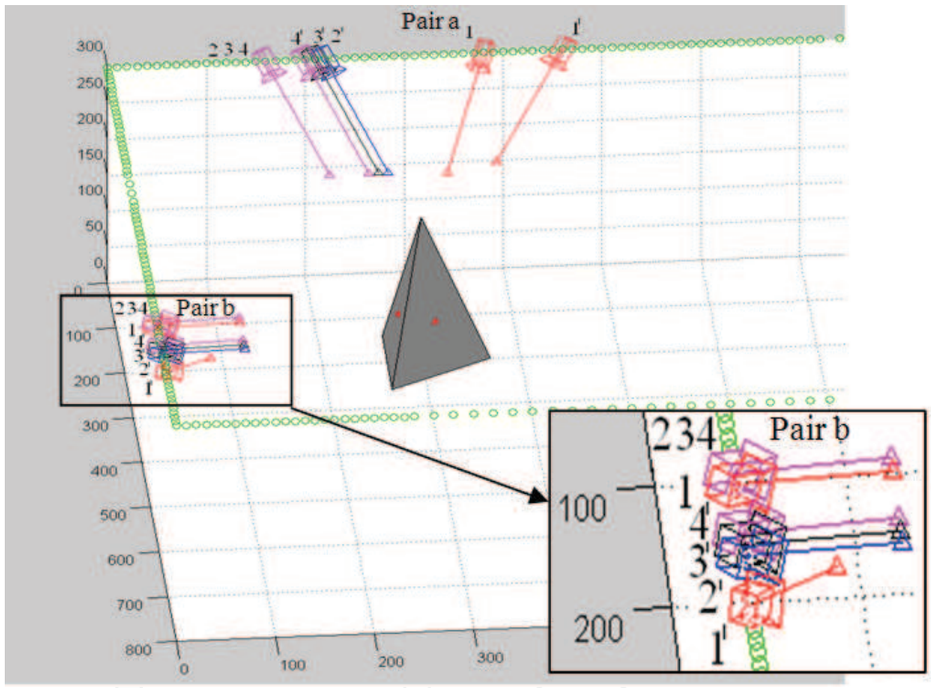


Fig. 14. The stereo pair positions and baseline lengths for the same target location vary according to the different accuracy.

IN	AF_{con}	B_a	B_b	Δ_{rec}	ΔZ_{max}
1	*8*	*100*	*80*	*3.3*	*4.5*
2	*11*	*70*	*60*	*4.5*	*6.2*
3	*14*	*60*	*50*	*4.9*	*7.9*
4	*17*	*50*	*40*	*5.2*	*9.7*

Table 1. The baseline lengths of two pairs and the reconstruction errors for different accuracy factors. *IN*: the index number; AF_{con}: the reconstruction accuracy requirement; B_a, B_b: baseline lengths for each pair, [cm]; Δ_{rec}: the maximum value reconstruction error of pairs, [cm]; ΔZ_{max}: the theoretical maximum depth error, [cm].

4. Conclusion

The chapter focuses on depth reconstruction, the arrangement of multiple stereo sensors for human activities space, which are important issues of IVAS. The research work can be concluded by means of two terms: the uncertainty analysis and human factor handling for vision system.

The analysis presented shows that the quantized 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 overly affect the depth accuracy, 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. By using the systems configuration and target properties, we can get reliable control over the shapes and intervals of the iso-disparity curves from the proposed iso-disparity mathematical model.

Modeling the target object as a tetrahedron gives a convenient way to extract the orientation of each surface and guarantee a good observability. Modelling the camera FoV using spherical co-ordinates simplifies the model and constraints, which speeds up computations. Formulating the stereo pairs with the 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 for tracking humans. In order to follow target objects movement, the camera movement distance constraints can be applied (Chen et al., 2007a).

The analysis of key factors which affect the accuracy of 3D reconstruction shows that the convergence angle and target distance are the 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 sampling of the cameras position has the flexibility to adjust the intervals and position ranges, and speed up computation.

5. References

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

Chakrabarty, K.; Iyengar, S; Qi, H. & Cho, E. (2002). Grid Coverage for Surveillance and Target Location in Distributed Sensor Networks. *IEEE Transaction on Computers*, vol. 51, no. 12, pp. 1448-1453, ISSN: 00189340.

Chen, J.; Khatibi, S. & Kulesza, W. (2007a). Planning of a Multi Stereo Visual Sensor System for a Human Activities Space. Proceedings of the 2nd International Conference on Computer Vision Theory and Applications, pp. 480 – 485, ISBN: 9789728865740 Barcelona, Spain, March 2007, INSTICC.

Chen, J.; Khatibi, S. & Kulesza, W. (2007b). Planning of a Multi Stereo Visual Sensor System - Depth Accuracy and Variable Baseline Approach. *Proceedings of IEEE Computer Society 3DTV-Con, the True Vision Capture, Transmission and Display of 3D Video*, ISBN: 9781424407224, Kos, Greece, May 2007, IEEE.

Chen, J.; Khatibi, S.; Wirandi, J. & Kulesza, W. (2007c). Planning of a Multiple Sensor System for Human Activities Space – Aspects of Iso-Disparity Surface. *Proceedings of SPIE on Optics and Photonics in Security and Defence*, vol. 6739, Florence, Italy, September, 2007, SPIE.

Chen, T.; Catrysse, P.; Gamal, A. & Wandell, B. (2000). How Small Should Pixel Size Be? *Proc. of SPIE on Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, vol. 3965, SPIE.

Chen, X. (2002). Design of Many-Camera Tracking Systems for Scalability and Efficient Resource Allocation. *PhD thesis*. ISBN: 0493628339, Stanford University.

Erdem, U. & Sclaroff, S. (2006). Automated Camera Layout to Satisfy Task-Specific and Floor Plan-Specific Coverage Requirements. *Computer Vision and Image Understanding*, vol. 103, pp. 156-169, ISSN 10773142.

Fusiello A.; Trucco E, and Verri A. (2000). A Compact Algorithm for Rectification of Stereo Pairs, *Machine Vision and Applications*, Vol.12, pp. 16-22, ISSN: 09328092.

Halif R. and Flusser J. (1998). "Numerically stable direct least squares fitting of ellipses," in: *Proc. of the 6th Int. Conf. Computer Graphics and Visualization*, pp. 125 – 132, Czech Republic, February, 1998.

Hashimoto, H. (2003). Intelligent Space: Interaction and Intelligence, *Artificial Life and Robotics*, vol. 7, no. 3, pp.79-85, ISSN: 14335298.

Hörster, E & Lienhart, R. (2006). On the Optimal Placement of Multiple Visual Sensors. *Proc. of ACM International Workshop on Video Surveillance & Sensor Networks*. ISBN: 1595934960, USA, October, 2006, ACM.

Mariottini, G.L. and Prattichizzo D. (2005). EGT for Multiple View Geometry and Visual Servoing Robotics Vision with Pinhole and Panoramic Cameras, *IEEE Robotics & Automation Magazine*, Vol. 12, No. 4, pp. 26-39, ISSN 10709932

MIT Project Oxygen, (2008). http://oxygen.csail.mit.edu/.

Mittal, A. (2006). Generalized Multi-Sensor Planning. *Proc. of 9th European Conference on Computer Vision*. ISBN: 978-3540338321, Austria, May, 2006, Springer.

O'Rourke, J. (1987). *Art Gallery Theorems and Algorithms*. ISBN 0195039653, Oxford University Press.

Ogle, K. (1950). Researches in Binocular Vision. W.B. Saunders Company, Philadelphia & London.

Pollefeys, M. & Sinha, S. (2004). Iso-disparity Surfaces for General Stereo Configurations. *Proc. of the 6th European Conf. on Computer Vision*. ISBN: 9783540219811, Czech Republic, May, 2004, Springer.

Samson, E.; Laurendeau, D.; Parizeau, M.; Comtois, S.; Allan, J. & Gosselin, C. (2006) The Agile Stereo Pair for Active Vision. *Machine Vision and Application*, vol. 17, no. 1, pp. 32-50.

Völpel, B. & Theimer, W. M. (1995). Localization Uncertainty in Area-Based Stereo Algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, no. 12, pp. 1628 – 1634, ISSN: 00189472.

Stereo Vision

Edited by Asim Bhatti

ISBN 978-953-7619-22-0

Hard cover, 372 pages

Publisher InTech

Published online 01, November, 2008

Published in print edition November, 2008

The book comprehensively covers almost all aspects of stereo vision. In addition reader can find topics from defining knowledge gaps to the state of the art algorithms as well as current application trends of stereo vision to the development of intelligent hardware modules and smart cameras. It would not be an exaggeration if this book is considered to be one of the most comprehensive books published in reference to the current research in the field of stereo vision. Research topics covered in this book makes it equally essential and important for students and early career researchers as well as senior academics linked with computer vision.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Wlodek Kulesza, Jiandan Chen and Siamak Khatibi (2008). Arrangement of a Multi Stereo Visual Sensor System for a Human Activities Space, Stereo Vision, Asim Bhatti (Ed.), ISBN: 978-953-7619-22-0, InTech, Available from: http://www.intechopen.com/books/stereo_vision/arrangement_of_a_multi_stereo_visual_sensor_system_for_a_human_activities_space

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen