Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Real-Time Obstacle Detection for an Autonomous Wheelchair Using Stereoscopic Cameras

Thanh H. Nguyen, Jordan S. Nguyen, Duc M. Pham, Hung T. Nguyen, *Senior Member, IEEE*

***Abstract*—This paper is concerned with the development of a real-time obstacle avoidance system for an autonomous wheelchair using stereoscopic cameras by severely disabled people. Based on the left and right images captured from stereoscopic cameras mounted on the wheelchair, the optimal disparity is computed using the Sum of Absolute Differences (SAD) correlation method. From this disparity, a 3D depth map is constructed based on a geometric projection algorithm. A 2D map converted from this 3D map can then be employed to provide an effective obstacle avoidance strategy for this wheelchair. Experiment results obtained in a practical environment show the effectiveness of this real-time implementation.**

I. INTRODUCTION

SEVERELY disabled people use power wheelchairs to assist their mobility and enhance their independence. Conventional power wheelchairs are not always sufficient to compensate for mobility disabilities. For many people, the operation of these wheelchairs is a difficult and demanding task. For this reason, the availability of smart wheelchairs is very appealing to this group of people [1, 2].

Cameras have been used on robots and power wheelchair to provide image data. They can be mounted on a power wheelchair in front of a user's face to detect head position and gaze direction to decipher his/her intended direction [3-5]. In order to obtain an obstacle map, several vision-based algorithms [6, 7] have been implemented to process images captured from cameras in literature.

Camera calibrations also play an important role for retrieving the 3D information of surrounding environments. Based on given constraints, self-calibration methods for moving stereoscopic cameras can be implemented effectively [8]. In addition, the two-dimensional correspondence search problem can be reduced to a scan-line search, greatly reducing both complexity and computational effort [9]. It is also possible to correct lens distorted images using the estimation of a lens distortion coefficient [10, 11].

This work was supported in part by Australian Research Council under Discovery Grant DP0666942 and LIEF Grant LE0668541.

Thanh H. Nguyen is with Faculty of Engineering, University of Technology, Sydney, Broadway, NSW 2007, Australia (phone: +612-9514-2451; fax: +61 2 9514 2868; e-mail: thnguyen@eng.uts.edu.au).

Jordan S. Nguyen is with Faculty of Engineering, University of Technology, Sydney, Broadway, NSW 2007, Australia (e-mail: Jordan.Nguyen@student.uts.edu.au).

Duc M. Pham is with Faculty of Engineering, University of Technology, Sydney, Broadway, NSW 2007, Australia (e-mail: Duc.M.Pham@student.uts.edu.au).

Hung T. Nguyen is with Faculty of Engineering, University of Technology, Sydney, Broadway, NSW 2007, Australia (e-mail: Hung.Nguyen@uts.edu.au).

The Sum of Absolute Differences (SAD) correlation method has been used for the detection of landmarks and corresponding images. It has also been used to match distances and estimation motion [12-14]. In particular, it has been applied to detect real-time automated landmarks in mobile robotic navigation and mapping [15].

Using stereoscopic cameras, in this paper we apply the SAD correlation method to estimate available free space to allow a power wheelchair to avoid obstacles in real-time. The paper is organised as follows. In Section 2, a brief background on stereoscopic vision is introduced. In Section 3, the computation of stereoscopic disparity from the left and right images is presented using the SAD correlation method. Section 4 shows the experiment result in which a 2D distance map has been converted from its 3D point map in the real-time environment. In Section 5, the real-time implementation based on camera calibration and stereo disparity is presented. Finally, Section 6 concludes the paper.

II. STEREOSCOPIC VISION

A. Camera Model

A pinhole image model is shown in Fig. 1. The image of a point P is the point $\mathbf{x} = [x, y]^T$ in the image plane through the optical centre [16]. The point $\mathbf{X} = [X, Y, Z]^T$ is projected onto the image plane at the point as

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} X \\ Y \end{bmatrix}, \tag{1}$$

The stereo camera model can be described through the below equations

$$\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \overbrace{\begin{bmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{bmatrix}}^{A} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{2}$$

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \overbrace{\begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}}^{D} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix}, \tag{3}$$

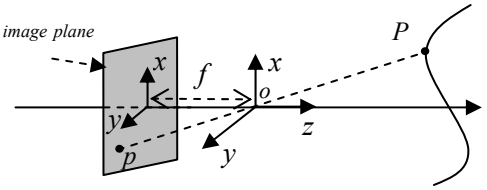

Fig. 1: Pinhole image model

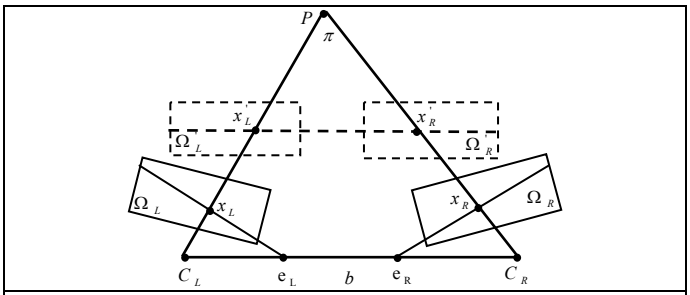

Fig. 2: Epipolar and rectified epipolar geometry

where the actual image coordinates are given by vector $\mathbf{x}' = (x', y', 1)^T$ instead of the ideal image coordinates $\mathbf{x} = (x, y, 1)^T$. Parameters of an additional transformation relative to parameters of the camera are s_x, s_y, s_θ and o_x, o_y. A is the intrinsic parameters, D is the extrinsic parameters which describes the rigid displacement between the world coordinates systems $\mathbf{X} = (X, Y, Z, 1)^T$ and camera coordinate system $\mathbf{X}_0 = (X_0, Y_0, Z_0, 1)^T$ by the rotation matrix R and translation vector T as in Eq. (3).

B. Epipolar Geometry

The image planes can be rectified to yield the horizontal image planes which are simpler for calculating disparity values as well as for determining the depth information.

The two cameras are indicated by their centres C_R and C_L and image planes Ω_L and Ω_R as shown in Fig. 2. Assume that P is a object point, the camera centres and its images x_R and x_L lie in a common plane π ($C_R P C_L$). The camera baseline b intersects each image plane at the epipoles e_L and e_R. Any plane π containing the baseline is an epipolar plane, and intersects the image planes Ω_L and Ω_R in corresponding epipolar lines $e_L x_L$ and $e_R x_R$. Given a unique geometry, the corresponding point x_L of any point x_R may be found along horizontal scan-lines corresponding to x_L' and x_R' on the image planes Ω_L' and Ω_R'.

III. Stereo Disparity

The purpose of stereo vision is to perform range measurements based on images obtained from stereoscopic cameras. Basically, there are two relevant steps:

- Establishing the correspondence between image features in different views of the scene.
- Calculating the relative displacement between feature coordinates in each image.

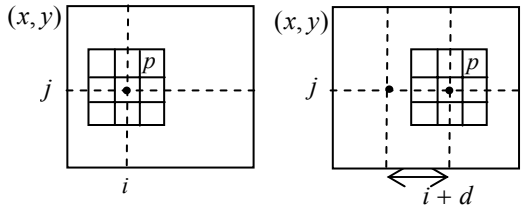

Fig. 3: Correlation of two 3x3 windows along corresponding epipolar lines

A. Establishing Correspondence

The main principle of correlation techniques is that a pixel in image 2 matches a corresponding pixel in image 1. In this project, the correlation mask is a square neighborhood around the pixel p as shown in Fig. 3.

In the SAD algorithm, the criterion for the best match is minimisation of the sum of the absolution differences of corresponding pixels in a window. In this project, the SAD correlation algorithm is implemented. In particular, for every pixel in the image, a neighborhood of a given square size is selected from the reference image. This neighborhood to a number of neighborhoods is compared in the other image (along the same row). Finally, the best match is selected and a disparity map is produced.

The comparison of a neighborhood is done using the sum of absolution differences as follows

$$SAD = \min_{d=d_{\min}}^{d_{\max}} \sum_{i=-M}^{+M} \sum_{j=-M}^{+M} \left| I_R(x+i, y+j) - I_L(x+i+d, y+j) \right|, \quad (4)$$

where the centre of a window of size $(2M+1)\times(2M+1)$ (called a correlation mask) is at the coordinates of the matching pixel (i, j), (x, y) are pixel coordinates in one image, I_L and I_R are the intensity functions of the right and left images and $d_{\min}$, $d_{\max}$ are the minimum and maximum disparities. The disparity $d_{\min}$ of zero pixels often corresponds to an infinitely far away object and the disparity $d_{\max}$ denotes the closest position of an object. If the disparity range is reduced, the system will run faster and will reduce the chance of mismatching.

B. Calculating Distances

The distances from the cameras are determined using the displacement between images and the geometry of the cameras. The position of the matched feature is a function of the displacement, the focal length of the lenses, resolution of the CCD and the displacement between cameras. In addition, equations of geometric projection are used to convert depth maps into distance images.

IV. Development of a 2-D Distance Map

A. Stereo Disparity Map

A power wheelchair has been fitted with stereoscopic cameras in the Key University Research Centre for Health Technologies as shown in Fig. 4. Fig. 5 and Fig. 6 show an image pair obtained from the stereoscopic cameras. Disparity is defined as the difference between the coordinates of the same features in the left and right images. Because the cameras are horizontally aligned, only the horizontal displacement is relevant.

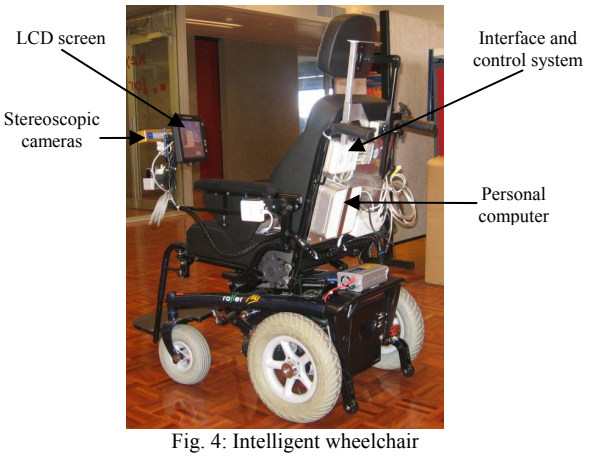

Fig. 4: Intelligent wheelchair

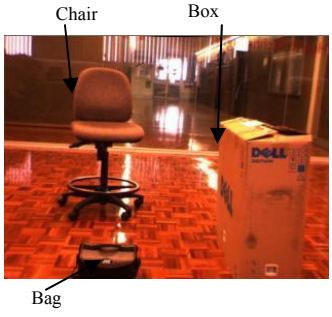

Fig. 5: The left image

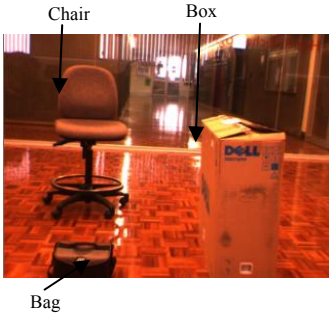

Fig. 6: The right image

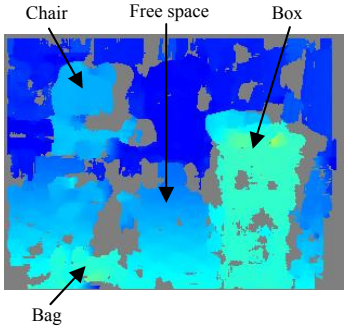

Fig. 7: Stereo disparity map

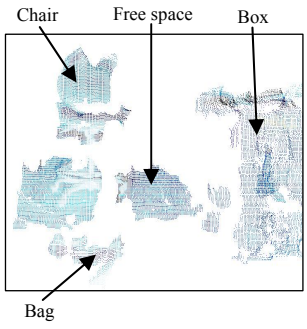

Fig. 8: 3D point map

The disparity map, as depicted in Fig. 7, using the SAD correlation method as shown in Eq. (4). This map displays the box, chair and bag which are considered as obstacles. The size of the mask used in this method must be proportional to the resolution of the images processed. Thus, in order to produce comparable results, we chose a 9x9 correlation mask for a 320x240 pixel image.

B. 3D Point Map

From the stereo disparity map, a 3D point cloud image is generated based on sub-pixel interpolation for more accurate stereo depth extraction. Each valid depth pixel is converted into a 3D XYZ position as shown in Fig. 8. In addition, the color of the point is also written out through the use of a rectified colour image.

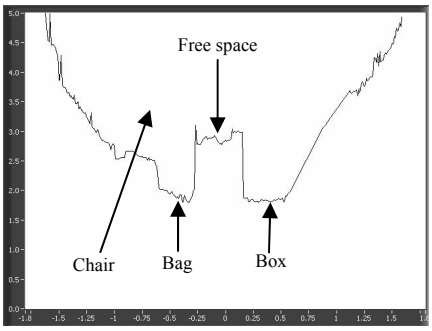

Fig. 9: 2D distance map

C. 2D Distance Map

From a 3D point image, properly there are three steps to convert into a 2D distance map as follows:

- Step 1: n 3D points (x_i, y_j, z_k) are converted into n 3D points (X_i, Y_j, Z_k) on coordinates X, Y, Z.
- Step 2: Assume that a certain value X_i on the horizontal X-axis corresponds to many points (Z_k, Y_j). The minimum distance $Z_{i\min}$ on Z-axis, from an object to the wheelchair corresponding to a point X_i on X-axis, is chosen.
- Step 3: In order to obtain a 2D distance map $(X_i, Z_{i\min})$, all minimum distances $Z_{i\min}$ corresponding to values X_i are computed as follows

$$Z_{i\min} = \min_{j=0}^{k} Z_{ij} . \quad (5)$$

The key problem of this paper is focused on determining a 2D map from a stereo disparity image based on vision methods in real-time. Distance values of this 2D map are very important for the implementation of obstacle avoidance and free space detection. For the example as described in this paper, the obstacles and free space are detected as shown in Fig. 9.

V. Real-time Implementation

An existing program, the Point-Grey Research (PGR) software, was used for distance detection from the stereoscopic images acquired by the Bumblebee camera. In this software, the offset values are used to generate a disparity image where there is a scale of colder colours such as dark blue representing objects further away to hotter colours such as red representing objects very close. The line of vision of the Bumblebee camera has a total range of 40.54°. This vision system produces a 15.03° maximum angle to the left and a 25.51° maximum angle to the right of the wheelchair due to the angle position of the camera.

The colour disparity image is separated into 38 columns of 8 pixels wide each and each segment corresponds to 1.08° of vision. Each segment is scanned and the warmest colour in each segment, registered as the closest part of any object, is saved into a colour box below that segment. The colour in each box is then converted into its Red, Green and Blue (RGB) components and the values for each of the 38 segments are saved in a table. These values will then be used

to determine the distance of the closest part of an object in each of the 38 segments. From calibration data, the relationship between the distance z and the RGB components is found to be:

$$z = 0.0011R^2 - 0.6727R + 220.36 \quad \text{for } z < 200cm \text{ , (6)}$$
$$z = -1.5389G + 586.81 \quad \text{for } z \geq 200cm \text{ . (7)}$$

The Wheelchair Control Program was implemented using the Bumblebee camera system. The PGR software was used for image acquisition and colour disparity extraction, Visual Basic 6.0 was used for the Wheelchair Control Program and NI USB-6008 was used for D/A conversion.

The implemented Wheelchair Control Program was programmed in Visual Basic 6.0. For this real-time implementation, at distances closer than 3m there is a maximum error of approximately $\pm 8cm$ and at distances further than 3m there is a maximum error of approximately $\pm 10cm$. The sample time for this real-time detection system is currently at 0.25 sec. An example of the obstacle mapping program with an obstacle standing at 1.8m distance is shown in Fig. 10.

The results of the autonomous wheelchair operation performance show that good detection of objects and effective avoidance method have been achieved. For this real-time implementation, the actual turning barrier gap is about $100cm \pm 20cm$ and a stop zone about $35cm \pm 15cm$.

VI. Conclusion

In this paper, the camera model, camera calibration and epipolar geometry are presented. A stereo disparity map using the SAD correlation method is also performed based on the left and right images. From this disparity map, a 3D point map is computed based on a geometric projection approach. In addition, a 2D distance map is generated from this 3D map. For real-time implementation of an autonomous wheelchair, a simpler variation of the 2D distance map is generated from the stereoscopic images. Experiment results on the power wheelchair in the practical environment have illustrated the effectiveness of the implementation.

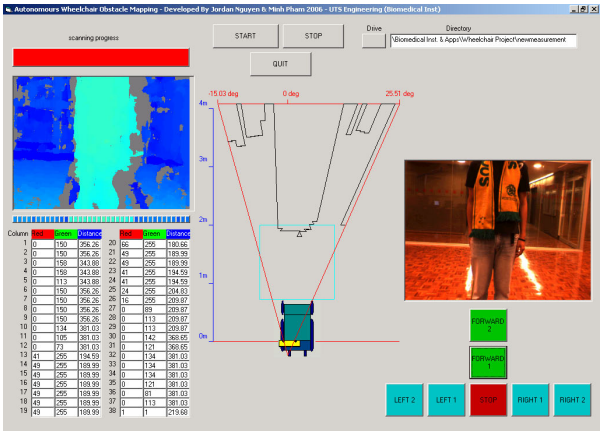

Fig. 10: Wheelchair control program

References

[1] R. Simpson, E. LoPresti, S. Hayashi, I. Nourbakhsh, and D. Miller, "The Smart Wheelchair Component System," *Journal of Rehabilitation Research & Development*, vol. 41, no. 38, pp. 429–442, 2004.

[2] R. C. Simpson, "Smart Wheelchairs: A Literature Review," *Journal of Rehabilitation Research & Development*, vol. 42, no. 4, pp. 423-436, 2005.

[3] Y. Adachi, K. Goto, Y. Matsnmoto, and T. Ogasawara, "Development of Control Assistant System for Robotic Wheelchair - Estimation of User's Behavior based on Measurements of Gaze and Environment," *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pp. 538-543, 2003.

[4] L. M. Bergasa, M. Mazo, A. Gardel, R. Barea, and L. Boquete, "Commands Generation by Face Movements Applied to The Guidance of A Wheelchair for Handicapped People," *Proceedings of the 15th International Conference on Pattern Recognition*, vol. 4, pp. 660-663, 2000.

[5] Y. Kuno, T. Yoshimura, M. Mitani, and A. Nakamura, "Robotic Wheelchair Looking at All People with Multiple Sensors," *Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Sys., pp*. 341-346, 2003.

[6] P. Saeedi, P. D. Lawrence, and D. G. Lowe, "Vision-Based 3-D Trajectory Tracking for Unknown Environments," *IEEE Transactions on Robotics and Automation*, vol. 22, no. 1, pp. 119-136, 2006.

[7] G. N. Desouza and A. C. Kak, "Vision for Mobile Robot Navigation: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 237-267, 2002.

[8] H. YU and Y. Wang, "An Improved Self-calibration Method for Active Stereo Camera," *Proceedings of the 6th World Congress on Intelligent Control and Automation*, pp. 5186-5190, 2006.

[9] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in Computational Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993-1008, 2003.

[10] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, "A Hough Transform-based method for Radial Lens Distortion Correction," *Proceedings of the 12th International Conference on Image Analysis and Processing*, vol. 2, pp. 182-187, 2003.

[11] W. Yu, "An Embedded Camera Lens Distortion Correction Method for Mobile Computing Applications," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 4, pp. 894-901, 2003.

[12] Y. W. Huang, S. Y. Chien, B. Y. Hsieh, and L. G. Chen, "Global Elimination Algorithm and Architecture Design for Fast Block Matching Motion Estimation," *IEEE Transactions On Circuits and Systems For Video Tech.* vol. 14, no. 6, pp. 898-907, 2004.

[13] J. Vanne, E. Aho, T. D. Hämäläinen, and K. Kuusilinna, "A High-Performance Sum of Absolute Difference Implementation for Motion Estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 876-883, 2006.

[14] K. Lengwehasatit and A. Ortega, "Probabilistic Partial-Distance Fast Matching Algorithms for Motion Estimation," *IEEE Transactions On Circuits For Video Technology*, vol. 11, no. 2, pp. 139-152, 2001.

[15] C. Watman, D. Austin, N. Barnest, G. Overett, and S. Thompson, "Fast Sum of Absolute Differences Visual Landmark Detector," *Proceedings of the 2004 iEEE International Conference on Robotics 8 Automation*, pp. 4827-4832, 2004.

[16] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision From Images to Geometric Models*. New York: Springer-Verlag, 2004.