Wearable Guide System for Obstacle Avoidance Based on Stereo Computer Vision

Jiexuan Feng[1, a *], Tong Wei[2,b]

[1 , 2]School of Instrumentation Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, China

[a]j.x.feng@163.com, [b]weitong@buaa.com

Keywords: computer vision, image processing, random sample consensus (RANSAC), blind guide, wearable

Abstract: ETA (Electronic Travel Aids) is an important method to guide the visual impaired to navigate, in which avoiding the obstacle effectively is the most essential content. A novel wearable ETA system and corresponding algorithms for obstacle avoidance based on binocular stereo computer vision were proposed in this paper. By utilizing RANSAC(Random Sample Consensus) algorithm, the equation of pavement plane is extracted from the images collected by the binocular cameras, and the obstacles are distinguished according to their height relative to the pavement plane, and obstacles coordinates relative to the blind in the world coordinate system are calculated and determined. This presented appliance can provide the information of walkable pavement and safe direction where there is no obstacle. The algorithms were verified by using the experimental prototype. The results show that it can detect obstacles effectively and stably with strong robustness, which provides good foundation for developing blind guide equipment further.

Introduction

The blind loses the visual sense because of the congenital or acquired physical defect, and then loses the ability to obtain knowledge and experience through the vision, so that they are very difficult in daily life, especially outgoing, in which avoiding obstacles is the most important aspect. It has been a priority research area to help the blind to avoid the obstacles effectively on the road. ETA (Electronic Travel Aids) is an important method to guide the visual impaired to navigate, in which how to avoid the obstacle effectively is the most essential content.

The traditional ETA instrument has to be held by hand, which greatly restricts the blind's hands activity, and causes a great deal inconvenience. The voice prompts are adopted usually in the traditional blind guide, and because of the environment noise, the blind cannot hear the voice guide clearly when he stays outdoor, which may cause some potential hazard. In order to make up these shortages, the wearable equipment have been the research mainstream nowadays for its more intuitive perception, more convenient carrying and better user experience.

Concerning obstacle detection, many non-vision methods, such as white cane, infrared detection, ultrasonic detection, and structured light detection[1,2] as well has been researched. However, the white cane's detecting range is too small, the infrared and ultrasonic appliance can only detect the obstacles with large volume, and the structured light detection method is too complex and expensive for the blind guide. In recent years, some obstacles detection methods based on computer vision have been presente. Some research institutes use monocular camera and SLAM(Simultaneous Localization and Mapping) algorithm to carry out the obstacle detection. However, this algorithm is very complicated for calculation, and the relationship with the physical scale cannot be established, and also it is very difficult to gain the coordinate of the obstacles[3, 4]. Because of the shortcomings of monocular vision, the ETA based on binocular computer vision has been developed in many universities. However, the presented methods to detect the obstacles by the binocular vision still exist some defects. Firstly, only the obstacles area in the image can be detected, but the position information of the obstacles relative to the blind in the real 3D space cannot be provided. Secondly, to extract the equation of pavement plane depends on the fixed relationship between the camera coordinate system and the world coordinate system, which is mainly used in the area of vehicle navigation, and is not suitable for ETA system to avoid the obstacles[5-7].

In view of the statement above, a novel wearable ETA system and corresponding algorithms for obstacle avoidance based on stereo computer vision has been designed and presented in this paper. Firstly through extracting the feature points from the image taken by binocular camera, the 3D points cloud in the field of vision is calculated. Then using RANSAC(Random Sample Consensus) algorithm, the equation of pavement plane is extracted. Finally the information of walkable pavement and safe direction where there is no obstacle can be provided. The algorithm is verified by using the experimental platform established for the wearable ETA system of avioding obstacle.

Experimental prototype introduction

The primary components of our system are shown in Fig.1. There are three parts in the experimental system: image collecting glasses, image processing host computer and intelligence vibratile vervel. Image collecting glasses is constituted by a binocular camera and a pair of glasses, where the binocular camera was set up on both side of the glasses by a bracket. The bracket can fix the relative position between the binocular camera and the glasses. The binocular camera was calibrated after it was installed on the glasses. Image processing host computer takes charge of binocular images processing, obstacle positon calculating, walking direction planning, and sending the directional signal to vibratile vervel. Vibratile vervel contains a slave computer and a couple of vibrators. The slave computer receives directional signal from host computer, translates it into frequency signal and drives vibrators with PWM wave. The two vibrators are fixed on both side of ankle, implying to turn left or right respectively, and the vibration frequency is proportional to the turning angle.

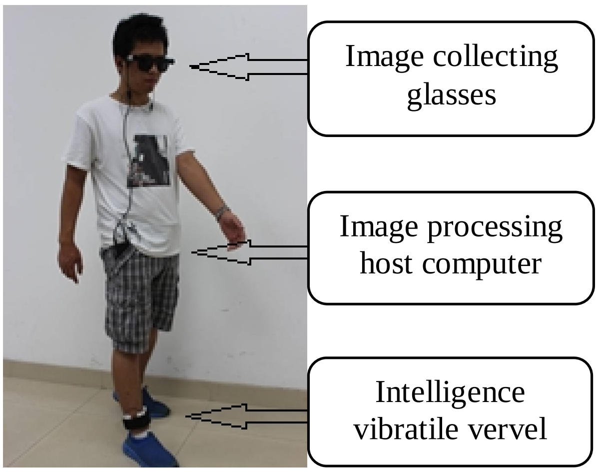


Fig.1 Wearable Guide System

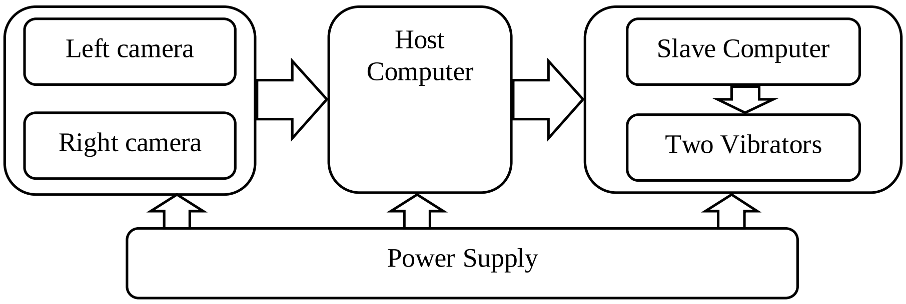


Fig.2 Block diagram of data flow

Obstacle avoidance algorithm

After the binocular images were transmitted to host computer, an image processing algorithm was adopted to process the images, which will be introduced as following.

3D points cloud extraction. Firstly, the feature points in the binocular images were extracted by SURF(Speeded Up Robust Features) algorithm. SURF algorithm takes advantage of rotation invariance, translation invariance and scale transformation invariance[8]. After using SURF algorithm to match the feature point sets, polar geometry was used to get rid of wrong matched couples, so that the reliable matched couples sets, P_l and P_r, are achieved. Then the following Eq.1 was used to calculate the 3D coordinates of the matched points in the camera coordinate system.

$$\begin{cases} x = \dfrac{z(u_l - C_{xl})}{f_{xl}} \\ y = \dfrac{z(v_l - C_{yl})}{f_{yl}} \\ z = \dfrac{f_{xl} f_{yl} (f_{xr} t_x + C_{xr} t_z - t_z u_r)}{f_{yl}(r_7 u_r - f_{xr} r_1 - C_{xr} r_7)(u_l - C_{xl}) + f_{xl}(r_8 u_r - f_{xr} r_2 - C_{xr} r_8)(v_l - C_{yl}) + f_{xl} f_{yl}(r_9 u_r - f_{xr} r_3 - C_{xr} r_9)} \end{cases} \tag{1}$$

In Eq.1, f_{xl} and f_{y} are the normalized focal length at the horizontal axis and vertical axis of left camera, f_{xr} and f_{yr} are the normalized focal length at the horizontal axis and vertical axis of right camera, (C_{xl}, C_{yl}) and (C_{xr}, C_{yr}) with the unit of pixel are the principal point coordinations of the left and right camera respectively. Let R_c and t_c be the rotation component and translation component of coordinate transformation matrix from the right camera to the left, r_1~r_9 the elements of R_c, t_x, t_y and t_z the elements of t_c, (x,y,z) (unit: mm) the coordinate in the camera coordinate system, (u_l,v_l) and (u_r,v_r) the coordinates of a pair of matched points in the image coordinate system.

Through the above calculation, the coordinates of 3D points cloud P in the current field of vision can be obtained.

The extraction of pavement plane based on RANSAC. There are two kinds of points in the 3D points cloud P, the points on the pavement plane and the ones on the obstacles. An algorithm named RANSAC can distinguish both of them and the pavement plane equation effectively[9]. The input of RANSAC algorithm is a group of observation data which can be parameterized into a model. The data matching the model are named as inliers, and the others are outliers. RANSAC algorithm can fit this parameterization model and find out the outliers from the input data by an iterative way. In this paper, input is 3D points cloud P, the outliers are the points on the obstacles, and the required model is the parameters of the pavement plane equation.

In order to extract the parameters of pavement plane equation by RANSAC, three unconditional and unco-linear points from 3D points cloud P are taken out randomly every time. Assuming they are inliers, least square method was used to fit the pavement plane equation $Ax+By+Cz=1$, and the parameterization model (A, B, C) was built. The next step is to use this model to test each of the rest of the 3D points. If one of the rest conforms the model, this point should be accepted as an inlier. For each of the points, the way to judge whether a 3D point p_i conforms the model is to calculate the distance from p_i to plane, being denoted as Di. If $Di \le D$mm, p_i should be made available for inliers set. After all of the rest has been tested, if the size of inliers set is bigger than d, the model (A, B, C) can be accepted as a reasonable model. At last, inlier sets are utilized to re-fit the plane equation. The error of plane model is introduced to evaluate the model, being denoted as e. e is the mean square deviation of D_i of inliers points, and the smaller the e, the more precise the plane model. An iteration time, denoted as k, is set to represent the cycle number of this program. The plane model which has the smallest e is selected as the best model, and should be recorded. The best inliers set should be record at the same time. In this paper, according to the demand of blind guide, set D=3mm, d=50% of input data size, k=35.

The algorithm can be described as follows:

(1) Initialize iteration time k according to the size of 3D points set, model error $e=\infty$, the size of inliers n=0;

(2) Take out three unconditional and unco-linear points from 3D points cloud P randomly, add these three points into inliers set, use square method to fit a plane equation $A_1x+B_1y+C_1z=1$;

(3) Use model (A_1, B_1, C_1) to test each of the rest of the 3D points, calculate distance from point to the plane Di, if $Di \le 3$mm, add the point into inliers set;

(4)Calculate the size of inliers set n', if n'>n, fit the new plane model and calculate the new model error e', if e'<e, update the plane model with (A_1, B_1, C_1), update n with n', update e with e';

(5)If the program has executed k times, accept the current model as the best model, then record the best inliers set and the outliers set, otherwise turn to step two.

Coordinate transformation for matched points collective. Both of the plane equation and the outliers of 3D points calculated in 3.2 section are based on the camera coordinate. The world coordinate $O_wX_wY_wZ_w$ is built to describe the position relation between human and the pavement plane. X_wY_w plane stands for the pavement plane. In order to build the connection between the camera coordinate and the world coordinate, coordinate transformation matrix is needed. As showed in Fig.3, $O_pX_pY_pZ_p$ is the camera coordinate, $O_wX_wY_wZ_w$ is the world coordinate. Y_w is the vertical projection on the pavement plane of Z_p, pointing at the walking direction. Z_w is perpendicular to the pavement plane. X_w is orthogonal to Y_w and Z_w. According to the pavement plane equation $Ax+By+Cz=1$which is based on the camera coordinate, the coordinate transformation matrix can be derived.

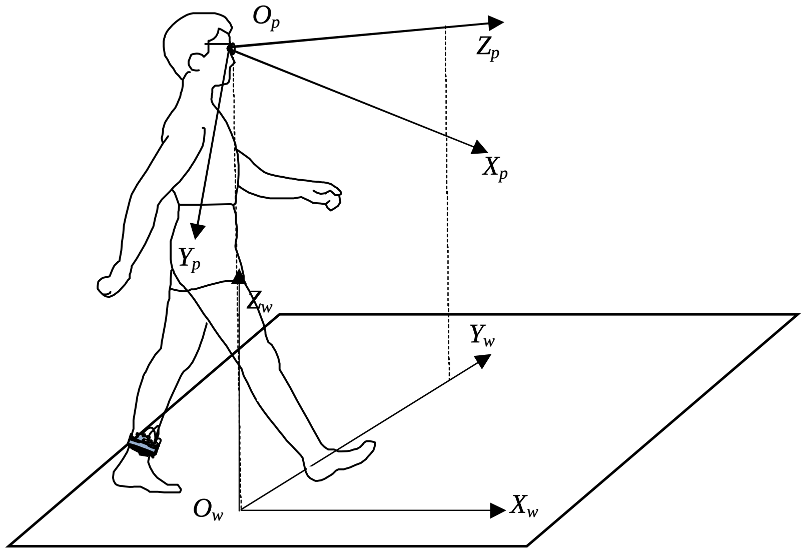


Fig.3 The relationship between the camera coordinate and the world coordinate

Y_w and Z_p can be expressed in camera coordinate as the vectors in Eq.2.

$$\begin{cases} \vec{Y}_w = \left[-\frac{A}{\sqrt{A^2+B^2+C^2}} \quad -\frac{B}{\sqrt{A^2+B^2+C^2}} \quad \frac{1}{C}-\frac{C}{\sqrt{A^2+B^2+C^2}} \right] \\ \vec{Z}_w = \left[-\frac{A}{\sqrt{A^2+B^2+C^2}} \quad -\frac{B}{\sqrt{A^2+B^2+C^2}} \quad -\frac{C}{\sqrt{A^2+B^2+C^2}} \right] \end{cases} \tag{2}$$

Thus the vector of X_w is shown in Eq.3.

$$\vec{X}_w = \vec{Y}_w \times \vec{Z}_w = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ -\frac{A}{\sqrt{A^2+B^2+C^2}} & -\frac{B}{\sqrt{A^2+B^2+C^2}} & \frac{1}{C}-\frac{C}{\sqrt{A^2+B^2+C^2}} \\ -\frac{A}{\sqrt{A^2+B^2+C^2}} & -\frac{B}{\sqrt{A^2+B^2+C^2}} & -\frac{C}{\sqrt{A^2+B^2+C^2}} \end{vmatrix}$$
$$= \begin{bmatrix} \frac{B}{C\sqrt{A^2+B^2+C^2}} & \frac{-A}{C\sqrt{A^2+B^2+C^2}} & 0 \end{bmatrix} \tag{3}$$

In order to make X_w parallel with X_p, the way to calculate the angle anticlockwise rotating around Z_p, θ degrees, is shown in Eq.4.

$$\theta = \begin{cases} \arccos\left(\frac{\vec{X}_w \cdot \vec{X}_p}{|\vec{X}_w||\vec{X}_p|}\right) = \arccos\frac{B}{\sqrt{A^2+B^2}} \left(A \ge 0\right) \\ -\arccos\left(\frac{\vec{X}_w \cdot \vec{X}_p}{|\vec{X}_w||\vec{X}_p|}\right) = -\arccos\frac{B}{\sqrt{A^2+B^2}} \left(A \le 0\right) \end{cases} \tag{4}$$

After X_w is parallel with X_p, in order to make Z_w parallel with Z_p, the way to calculate the angle anticlockwise rotating around X_p, ψ degrees, is shown in Eq.5.

$$\psi = \arctan\left(\frac{\frac{1}{\sqrt{A^2+B^2+C^2}}}{\frac{1}{C}}\right) + \frac{\pi}{2} = \arctan\frac{C}{\sqrt{A^2+B^2+C^2}} + \frac{\pi}{2} \tag{5}$$

Thus, the rotation matrix $\boldsymbol{R}$ and translation matrix $\boldsymbol{t}$ between camera coordinate and world coordinate can be expressed as Eq.6 and Eq.7 respectively.

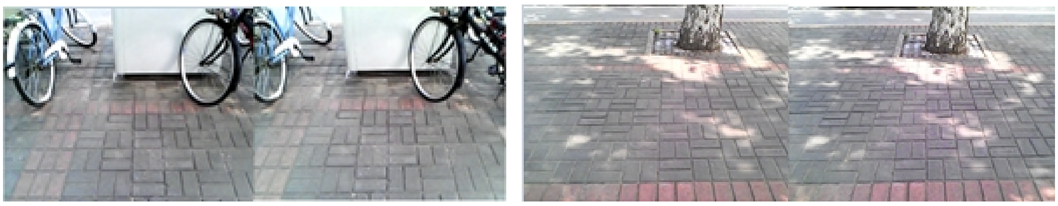
$$\boldsymbol{R} = \boldsymbol{R}_{Xp}(\psi)\boldsymbol{R}_{Zp}(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ \cos\psi\sin\theta & \cos\psi\cos\theta & \sin\psi \\ \sin\psi\sin\theta & \sin\psi\cos\theta & \cos\psi \end{bmatrix} \tag{6}$$

$$\boldsymbol{t} = \begin{bmatrix} \frac{A}{\sqrt{A^2+B^2+C^2}} \\ \frac{B}{\sqrt{A^2+B^2+C^2}} \\ \frac{C}{\sqrt{A^2+B^2+C^2}} \end{bmatrix} \tag{7}$$

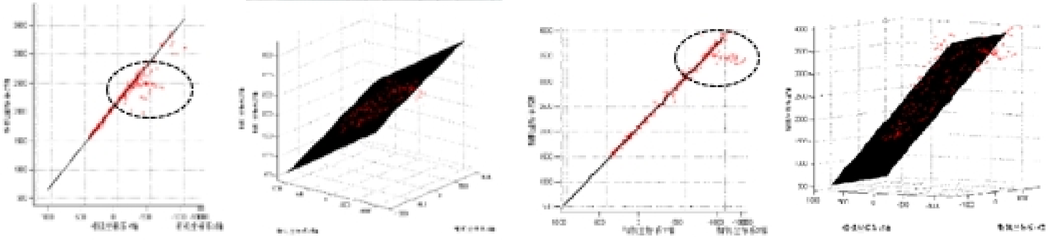
$$\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix} \begin{bmatrix} x_p \\ y_p \\ z_p \\ 1 \end{bmatrix} \tag{8}$$

By using Eq.8, 3D points on the obstacles can be transformed from camera coordinate to world coordinate. The distribution of obstacles in the view should be available. After the distribution is transported to slave computer, it would turned into a frequency signal based on PID control algorithm and control two vibrators by PWM wave.

Verifying experiments for abstacle avoidance algorithm. In order to verify the validity of the proposed algorithm, a binocular camera was used in this paper to sample some outdoor scenes randomly for the experiments. Fig. 3 is two groups of binocular images with obstacles in the field of vision. In Fig. 4(a), a total of 267 matched feature points are acquired by using SURF algorithm. Then the 3D points cloud was obtained by the theoretical calculations and the Matlab was used to draw the 3D coordinate system. After screened by RANSAC algorithm, the pavement plane equation and the 3D points on the pavement were obtained. The number of points on the pavement is 186, the number of points on the obstacle is 81, and the percentage of the outliers is 30.3%. Then the pavement plane is drawn with the fitted plane equation with Matlab. From the Fig. 4(a)-3, we can see that the points of this plane and the points on the surface of the pavement are basically coincident. In Fig.4 (b), a total of 328 3D points are acquired, where the points on the pavement is 211, that on the obstacle is 117, and the percentage of the outliers is 35.6%. The Fig. 43(b)-3 shows that the pavement plane fitting is still correct.



(a)-1 Left image (a)-2 Right image (b)-1 Left image (b)-2 Right image



(a)-3 Position of plane and obstacle (b)-3 Position of plane and obstacle

Fig. 4 3D points cloud and the plane model positon pictured by Matlab

Table. 1 shows a measure result when outliers proportion is larger than 30%. From Table.1, by using the algorithm in this paper, the absolute error in the measured angle is lower than 2 degrees, and the relative error is lower than 5%. Experiments show that this precision meets accuracy requirements of blind guide.

Table.1 Angle between the actual ground and measured plane(Outliers Proportion>30%)

	Actual Angle[°]	Measured Angle[°]	Absolute Error[°]	Relative Error
X Label	-25	-23.942	1.048	4.1%
	-10	-9.676	0.324	3.2%
	10	10.340	0.340	3.4%
	25	26.209	1.209	4.8%
Z Label	20	20.221	0.221	1.1%
	30	31.410	1.410	4.7%
	40	38.185	1.815	4.5%

Conclusions

A novel wearable ETA system for obstacle avoidance based on computer binocular stereo is presented in this paper. Firstly, the matched feature points' 3D coordinate under the camera coordinate system is extracted by binocular stereo vision algorithm, forming 3D points cloud. Then the pavement plane is extracted from points cloud by RANSAC algorithm. Accordingly, the points on obstacles are separated, and the pavement plane equation can be obtained. At last, the conversion relationship between the camera coordinate and world coordinate is calculated by the pavement plane equation, obtaining the obstacle's position relative to the people's location in the world coordinate. That is to say, the obstacles' distribution situation is obtained to complete the guide work successfully.

By using guide vision system experimental platform for obstacle avoidance to verify the algorithm, the results show that the presented algorithm can calculate the pavement plane equation effectively and can be steady to distinguish the obstacle area robustly, so that it can provide obstacles' location information effectively for helping blind people navigation.

References:

[1] Sachithra Hemachandra,Thomas Kollar, Nicholas Roy and Seth Teller. Following and Interrupting Narrated Guided Tours[J].2011

[2] Walter M R, Hemachandra S, Homberg B, et al. Learning semantic maps from natural language descriptions[C]. Robotics: Science and Systems, 2013

[3] Roberto Manduchi, James M. Coughlan. The Last Meter: Blind Visual Guidance to a Target[J].Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.2014.04

[4] O. Gallo, M. Tico, R. Manduchi. Metering for Exposure Stacks[J]. EUROGRAPHICS 2012 / P. Cignoni, T. Ertl,2012,31(2)

[5] Vivek Pradeep, Gerard Medioni and James Weiland. A Wearable System for the Visually Impaired[J]. 32nd Annual International Conference of the IEEE EMBS.2010.08

[6] Pradeep V, Lim J. Egomotion estimation using assorted features[J]. International journal of computer vision, 2012, 98(2): 202-216

[7] Pradeep V, Medioni G, Weiland J. Robot vision for the visually impaired[C]//Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. IEEE, 2010: 15-22

[8] Zhang Z. A flexible new technique for camera calibration[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000, 22(11): 1330-1334

[9] M. A. Fischler and R. C. Bolles. Ramdom sample consensus: A paradigm for model fitting with applicantions to image analysis and automated catography. IJCV, 22(2): 125-140, 1997