Coordination of SLAM and Artificial Landmark Recognition using 2D/3D Sensors for Mobile Robot Vision

C. Joochim[1), C. Netramai[2), H. Roth[1)

1) Institute of Automatic Control Engineering, Hoelderlinstr3, 57068.
{chanin.joochim,hubert.roth}@uni-siegen.de
2) Center for Sensor System (ZESS), Paul-Bonatz-Strasse9-11, 57076.
netramai@zess.uni-siegen.de, University of Siegen, Germany.

Abstract: *Besides the high performance simultaneous localization and mapping algorithm (SLAM) has to build the 3D mapping at the same time and to estimate the pose of the locomotion in real time, the flawlessly mobile robotic application tasks, meanwhile should be able to recognize objects in the environment in order to achieve the variety practical missions. This paper hence presents a coordination between the real-time SLAM and artificial landmark recognition by fusion data from the 2D/3D sensors. The new 3D sensor namely the Photonic Mixer Devices (PMD) purposes real-time capturing the surrounding volume. The 2D high resolution image is registered on the 3D volume subsequently rescaling and calibration both sensors. Visual input from the 2D camera not only delivers high resolution texture data on 3D volume but also use for object recognition. Moreover, the ICP algorithm is taken over the theatrical image registration due to yield the real time data frames registration.*

***Keywords*:** PMD camera, SLAM, ICP algorithm, mobile robotics and image registration.

1. INTRODUCTION

Nowadays, the complex functions are included in one compact mobile robotics such as a wireless/GPS communications, smart obstacle avoidances, vision system as well as self localization techniques. The on-line, real time three dimension map building is also a very challenging task for modern mobile robotic systems to obtain the realistic visual appearance of particular environmental volume. The simultaneous localization and mapping (SLAM) is a technique used for map building. When a mobile robotic moves through an unknown environment, can localize themselves autonomously.

The SLAM can be represented by various active sources e.g. ultrasonic sensors and laser sensors, or by passive sources, i.e. vision sensors, from many kinds of cameras. Such cameras can be used to find unique characteristics of features based on the pixels in and around the features, whereas the active sources can't. Additionally, laser scanners have been widely used in mobile robotics because they can ideally give the precision of raw data. However, some drawbacks still have to improve, for example an inferior output in cluttered environments, high cost, heavy and consumable energy. One outstanding sensor, namely PMD camera, which can provide 3D data without robust against ambient light and weakly textured area, dominated in 3D scene applications [1][2], is proposed here. It is capable of capturing reliable depth images directly in real-time. The PMD is also compact and affordable, which makes it attractive for versatile applications including surveillance and computer vision. Furthermore, a resulting gray scale image can be applied for basic vision recognition. For these reason, PMD is a novel attractive tools for implement the SLAM.

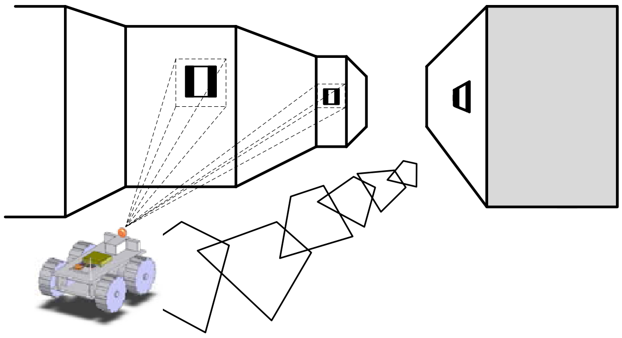


Fig.1-Coordination of SLAM and landmark recognition

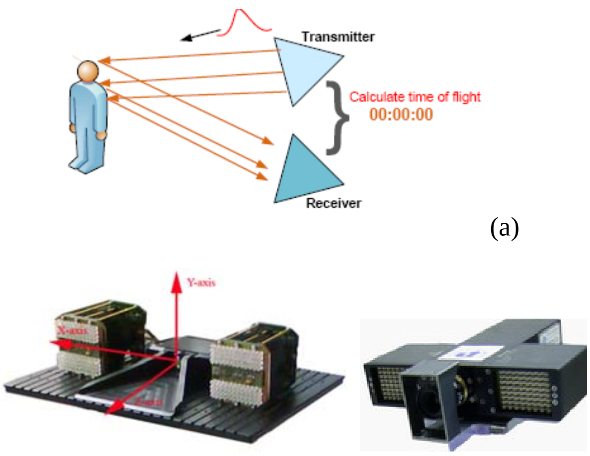
However, the high performance of SLAM is not only building a 3D mapping in real time but also should recognize objects, avoiding the obstacle as well as estimate the trajectory simultaneously. The prominent output from PMD is an each pixel depth measurement. The high resolution 2D camera is then combined in order to compensate the lack of complicated image processing and support the complex machine learning requirements. Nevertheless, searching the objects within a complex environment isn't the easy tasks. The Haar-like Features is one method, which the searching is take place in real-time. This method can provide accuracy and robustness for detecting the objects. It can be also applied to detect any features, patterns, shapes, color even in complex environmental. Figure 1 demonstrates an over all idea for the coordination SLAM and artificial landmark recognition. Moreover, the differential pixel resolutions and position overlapping due to machine setting up are considered in this work. The MERLIN (Mobile Experimental Robots for Locomotion and Intelligent Navigation) is used in a test platform. This robot is the high performance indoor environment. It communicates to a work station via a wireless module, and can be automatically controlled, or can be manually controlled by a joystick.

The Iterative Closest Point (ICP) is implemented for 3D map building. The ICP is based on searching of the nearest point-to-point, point-to-tangent plane pairs and point-to-projection, and additionally estimating the rigid transformation, which aligns them. The main of arduous computing part of ICP is an exhaustive search for correspondence.

For the comprehensive results, this paper augments the SLAM using the novel alternative output after combination a high resolution 2D camera with depth data from PMD. The brief summary principles of the PMD camera, data fusion, images registration as well as future suggestion plan are presented in this paper.

2. THE PRINCIPLE OF PMD CAMERA

The so called PMD camera as a smart pixel can be an integration element for a 3D imaging camera on a chip based on standard CCD- or CMOS-technology [3]. The main component is an array sensor. It can measure the distance to the target in parallel without scanning. The key execution is based on Time-Of-Flight principle. A light pulse is transmitted from a sender unit, and the target distance is measured by determining the turn around time back to the receiver. According to the speed of light, the interval distance can easily be calculated. Figure 2 shows models and the principle Time of Flight based on PMD camera.



(a)

(b) (c)

Fig.2 - a) Time of Flight principle PMD cameras b) A2 c) 13k

The PMD chip is the prominent component, which provides depth information in each pixel of the corresponding point in the object plane. Additionally, the PMD camera has the advantage of fast image mapping. This camera enables fast optical sensing and demodulation of incoherent light signals in one component. It also gives both intensity and distance for each pixel. The PMD can be used to get the excellent depth information as well as gray scale value of the scene. Currently, the PMD sensor devices provide the resolutions of pixels 48*x*64 pixels (model 3k), 64*x*16 pixels (model A2) and 160*x*120 pixels (model 19k). A common modulation frequency is 20 MHz, which results in an unequivocal distance range of 7.5 to 40 meters. The PMD sensor calculates the distance between obstacle and camera by measuring the phase shift. The depth data is obtained from the phase shift of the out-coming and incoming signals. The equation for autocorrelation is:

$$c(\tau) = \int_0^T x(t)x(t-\tau)dt \qquad (1)$$

Where T is the time of integration, the correlation is implemented by using four samples, $C_1 ... C_4$, with the time interval of $T/4$ and phase shift (ϕ).

$$\phi = \arctan(\frac{C_1 - C_3}{C_2 - C_4}) \qquad (2)$$

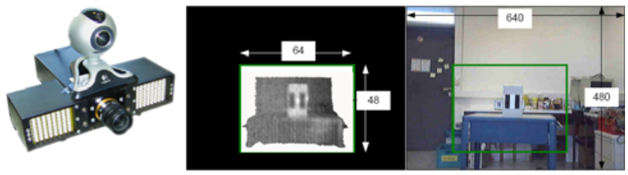
The distance (d) can be easily calculated by:

$$d = \frac{c_0 . \phi}{4\pi . f_{\text{mod}}} \qquad (3)$$

Where c_0 is the speed of light and f_{mod} is the modulation frequency.

3. CAMERAS SETUP AND CALIBRATION

The PMD camera that used in the experiment is 3k-S models with the resolution 64×48 pixels, horizontal field of view (FOV) is approximately 10.0° and 12.5° from vertical. The 2D camera has the resolution of 640×480 pixels, 45° and 34° from horizontal and vertical, respectively. The differential FOV and equipping position between both cameras impact on the overlap between output frames. Figure 3(a-c) shows the differential field of view from PMD and 2D cameras.



(a) (b) (c)

Fig.3 - a) Fusion 2D/3D; Output b) Depth PMD c) 2D

Hence, in this study, the calibration method is considered. The Camera calibration is an essential step before handle image processing tasks in order to extract metric information from image frames, especially the calculation between two cameras. This research acquires image data from two camera types, PMD and 2D camera. The camera calibration is therefore the most essential procedure before going forward in the next step. [4][5] proposed a flexible technique for calibration, which requires the camera to observe a planar pattern shown at a few different orientations. However, the population calibration tools by using OpenCV library base on [6] method is used in this research. The algorithm implements in camera calibration toolbox for MATLAB and OpenCV Intel C++ library.

4. AN IMAGE REGISTRATION

Overall steps of SLAM and object recognition are to integrate many algorithms in order to get a final outcomes. Figure 4 shows the algorithm box sets. The 2D camera uses for demanding and controlling the behavior of image registration. Thus, many algorithms are applied before yielding the final output. The PMD camera is used for providing the exact depth data to generate the high performance SLAM in real-time. The detail of every box set is proposed in following sections.

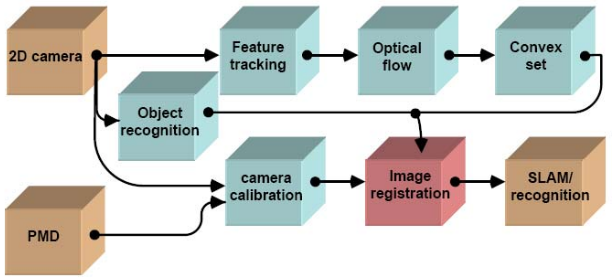


Fig.4-Diagram of 3D mapping generation.

4.1 GOOD FEATURE TRACKING AND CORRESPONDENCE FINDING

In the first step for SLAM, it is assumed that the mobile robotic is relocated from the first position to another. The correspondence points between two frames have to be found in order to merge those frames. The good feature tracking is a feature point extracted from an image. The Open Source Computer Vision Library, OpenCV is the very famous algorithms for image processing and computer vision. Some libraries are used in basic image analysis such as corner detection, canny edge detection and non-maxima suppression[7]. They proposed an iterative image registration technique for the stereo vision application, which is the well-known good feature tracking. This technique computes the flow for each pixel between image frames. Figure 5 illustrates the corresponding point example between two image frames.

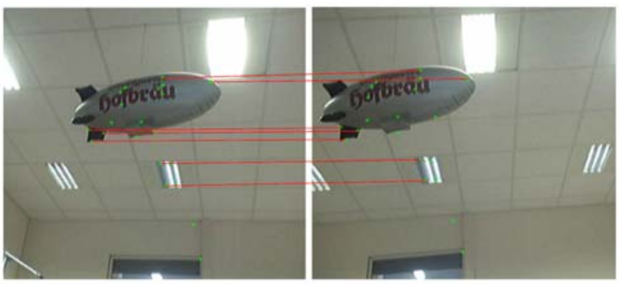


Fig.5 - Corresponding points two image frames

4.2 OPTICAL FLOW

The unpredictable mobile robot locomotion for obstacle avoidance is a complex task to generate 3D mapping. In addition, posing robot, for instance turn left, right, forward, backward and sloppy climbing up is an arduous achievement in the 3D mapping. In fact, the direction that should merge the frame is very importance. The optical flow is hence perused to seek the pose estimation of mobile robotic. One of the famous optical flow technique is the motion template. It was proposed by [8][9]. They proposed effective methods to track the object movement using the differencing edge of segmented moving of silhouettes for camera from frame-to-frame. A model of zeppelin airship movement is demonstrated in figure 6. The white indication is the current zeppelin position. Coming up of new silhouettes are captured in next frame and overlaid on the white position. Each step of fading sequences of silhouettes are recorded and referred to as the motion history image (MHI).

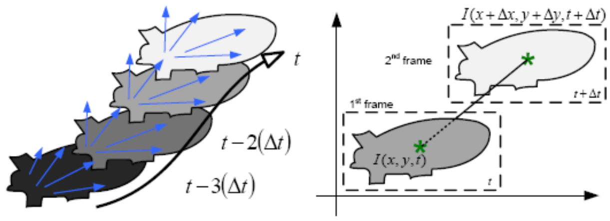


Fig.6-Optical flow of motion

The image input of the camera frame should sufficiently contain texture information in order to estimate the correspondence points between two frames. It is assumed that in figure 6, the first frame is captured at time t and point (x, y). This point contains the color intensity $I(x, y, t)$. When the zeppelin changes the position, meanwhile camera capture the second frame but the same point is still keep in the second frame. The intensity in the present frame is the same as in the previous frame.

$$I_2(x+\Delta x, y+\Delta y, t+\Delta t) = I_1(x, z, t) \tag{4}$$

Where x, y is the coordinate in x-axis and y-axis, respectively. The Tayloy serious represents the equation 4 as a summation term.

$$I_2(x+\Delta x, y+\Delta y, t+\Delta t) = \tag{5}$$
$$I_1(x,z,t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t + ...$$

It is assumed that the higher order terms are very small and can be ignored. Equation 5 is equal to equation 6.

$$I_1(x,z,t) = I_1(x,z,t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t \tag{6}$$

$$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = 0 \tag{7}$$
$$\frac{\partial I}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t}\frac{\Delta t}{\Delta t} = 0$$

$$\frac{\partial I}{\partial x}v_x + \frac{\partial I}{\partial y}v_y + \frac{\partial I}{\partial t} = 0 \tag{8}$$

Where $v_x = \frac{\Delta x}{\Delta t}$ and $v_y = \frac{\Delta y}{\Delta t}$ are the image velocity or optical flow at pixel (x, y). The $\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}$ and $\frac{\partial I}{\partial t}$ are the point intensity gradient. This equation can be rearranged concisely.

$$I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y}, I_t = \frac{\partial I}{\partial t} \quad (9)$$

$$I_x V_x + I_y V_y = -I_t$$

$$\nabla I . \vec{v} = -I_t \quad (10)$$

Where ∇I is the intensity gradient and $\vec{v}$ is the image velocity at pixel (x, y).

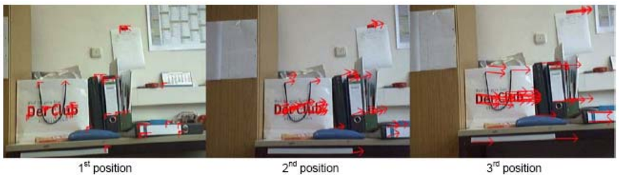


Fig.7 - Optical flow output

Figure 7 shows the optical flow output. The camera moves from right side to left side with speed approximately 10cm/1sec. Moreover, the arrow length expresses the direction of mobile robot locomotion. The directions of optical flow are used to find the direction for merging the image frames in order to generate SLAM in real time.

4.3 CONVEX SET

After realizing the direction of robot, the next issue is how to find the overlapping of image frames. It can be noticed from figure 9 that the mobile robot moves from the previous position (1st) to the next position (2nd). Some corresponding points still appear on both frames. To determine this, the fixing boundary of corresponding points is essentially to set up. The Quick hull, Graham's scan, Jarvis March and Convex hull are very famous for determining the smallest interested convex of clown points. Figure 8 demonstrates the steps to find the convex hull. Step1: finding the lowest point (L) within all set points. Step2: sorting set points in counterclockwise direction and network for each point. Step3: sorting by calculating the relative angle. The orientation of three points p, q and r are the ordering points in network. The angles ϕ_{pq} and ϕ_{pr} in counterclockwise are determined. The smaller angle is a convex vertex; removed a nonconvex from the network.

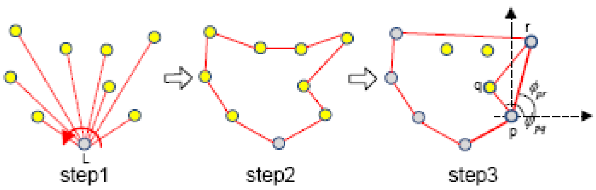


Fig.8 - Computation of Convex Hull

Figure 9 shows the boundary of corresponding area when camera changes the positions.

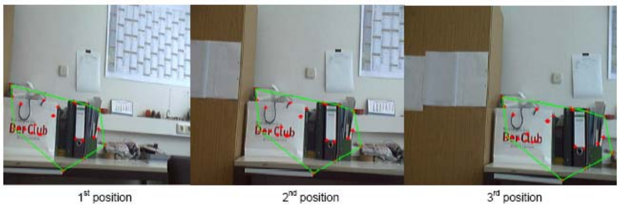


Fig.9 - Corresponding area

4.4 LANDMARK RECOGNITION

This work adapts the Haar-like Features with OpenCv library [10][11]. The Haar-like Features is the real-time, accuracy and robustness approach for detecting the objects. It can be applied to detect their features, patterns, shape and color even in complex environment.

Fig.10 - Positive sample in every angle of artificial landmarks used for training classifiers

The Cascade of Classifiers has to collect the negative and positive samples. The negative samples are the images that must not contain any object inside for detecting requirement, artificial landmark. In this experiment, the 7170 negative samples are used. The positive samples are the images that must contain the artificial landmark, and there are 1500 positive samples. Figure 10 shows the positive samples for the classifiers. This work uses the Intel Pentium 4 1.8GHz, 3GB RAM. The training times are approximately 80 hours. Figure 11 shows the output to detect artificial landmark in real time

Fig.11 – Output of recognition artificial landmark

5. IMAGES REGISTRATION

The 3D mapping acquires pose estimation of mobile robot. The 3D geometry is simultaneously collected by two sources while robot moves through the captured scene. The main procedure to grab three dimension screens is presciently robot pose estimation. Then, the 3D mapping yields the texture for the 3D model. If the mobile robot is able to precise locomotion, the registration 3D mapping can be effortlessly generated. Actually, unpredictably noise, relative pose as well as bumpy trajectory that cause knottiness the geometric overlapping for capturing scenes has to be considered for image registration.

In general, the iterative closest point (ICP) is broadly used for registration of 3D clouds. [12] presented the similar transformation parameters in m-dimensional space that give the least mean squared error between these point patterns and introduceed to solve the special absolute orientation problem by erecting a reduced gradient algorithm together with its convergence proof and by generalizing it to the case with weighted errors. This is based on the classic absolute orientation technique. An applied ICP algorithm justifie to generate mapping. The results are very excellent when apply to solve SLAM problem. The general solutions of mapping are used to find the similarity transformation parameters. These give the minimum value of the mean squared error of two point patterns. It is assumed that two different partial image patterns, $\mathbf{P} = (p_1, p_2, ..., p_n)$ and $\mathbf{Q} = (q_1, q_2, ..., q_n)$ are cloud interaction points between two robot positions. The aim is to properly estimate the next position for getting an ideal minimum error.

$$\mathbf{P}(p_1,..., p_n) - \mathbf{E}(e_1,..., e_n) = \mathbf{Q} = (q_1,..., q_n) \quad (11)$$

$$E(\mathbf{R}, \mathbf{t}, c) = \frac{1}{n}\sum_{i=1}^{n} \left\| p_i - (c\mathbf{R}q_i + \mathbf{t}) \right\|^2 \quad (12)$$

Where $\mathbf{E}(e_1,...,e_n)$ is an error vector. Equation (12) is an observation equation, which is achieved by least squares minimization and $(\mathbf{R}, \mathbf{t}, c)$ represents the optimum rotation, translation and scaling. Where $U\Lambda V^T$ is a singular value decomposition of PQ^T, $(UU^T = VV^T = I, D = diag(d_i), d_1 \geq d_2 \geq ... \geq d_n \geq 0)$

$$\mu_p = \frac{1}{n}\sum_{i=1}^{n} p_i \quad (13)$$

$$\mu_q = \frac{1}{n}\sum_{i=1}^{n} q_i \quad (14)$$

$$\sigma_p^2 = \frac{1}{n}\sum_{i=1}^{n} \left\| p_i - \mu_p \right\|^2 \quad (15)$$

$$\sigma_q^2 = \frac{1}{n}\sum_{i=1}^{n} \left\| q_i - \mu_q \right\|^2 \quad (16)$$

$$\Sigma_{ab} = \frac{1}{n}\sum_{i=1}^{n} (q_i - \mu_q)(p_i - \mu_p)^T \quad (17)$$

Where Σ_{ab} is a covariance matrix of $\mathbf{P}$ and $\mathbf{Q}$, while μ_p and μ_q are mean vectors of $\mathbf{P}$ and $\mathbf{Q}$. σ_p^2, and μ_q are variances around the mean vectors of $\mathbf{P}$ and $\mathbf{Q}$ respectively.

$$E(\mathbf{R}, \mathbf{t}, c) = \sigma_q^2 - \frac{tr(DS)^2}{\sigma_p^2} \quad (18)$$

$$\Lambda = I \quad if \ \det(\sum\nolimits_{pq} \geq 0) \quad (19)$$

$$\mathbf{R} = U\Lambda V^T \quad (20)$$

The registration can be used to calculate the optimal rotation by (20). The effect of rotation matrix can be immediately solved the translation as in equation (21).

$$\mathbf{t} = c_q - \mathbf{R}c_p \quad (21)$$

$$c = \frac{1}{\sigma_p^2} tr(\mathbf{DS}) \quad (22)$$

6. SIMULATION RESULTS

In order to prove the previous theory, the simulation results using MATLAB is discussed in this section. The simulation presents two crowning points with differential rotation and translation element. From the figure 12, $x = \sin(t), y = \cos(t), z = t$ are assumed as the first data set, where t is a various time. The data1 is multiplied with random rotation matrix, R, and translation matrix, T. The ICP algorithm calculates the translation matrix, rotation matrix and both registration data as shown in figure 12 (b). The figures 12(c) and (d) show the random data using a pseudo-random generated function. The simulation results prove that the translation and rotation matrix are found and matched with differential two corresponding points. The simulation results give the appropriate results can be applied in the next step of experiments.

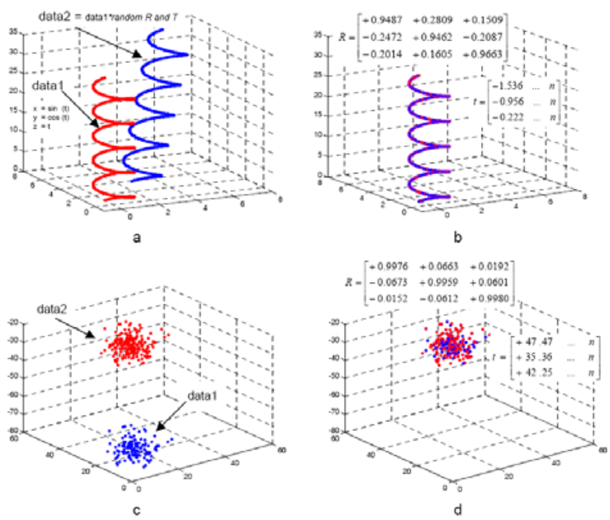
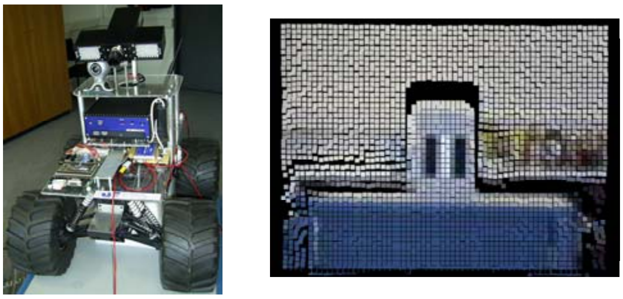


Fig.12 - Simulation results of rotation and translation matrix using ICP

7. EXPERIMENTAL RESULTS

In the experiment, the mobile robot MERLIN (Mobile Experimental Robots for Locomotion and Intelligent Navigation) is used as a test platform. It can be used for high performance indoor and outdoor off-road . The robot is equipped with a PMD, a 2D camera, a 16bit microcontroller and an embedded PC. The PMD and the 2D cameras are mounted on the top with a proper inclination angle. Figure 13 shows the mobile robot MERLIN and (b) shows the one frame output from combination between 2D/3D sensors.

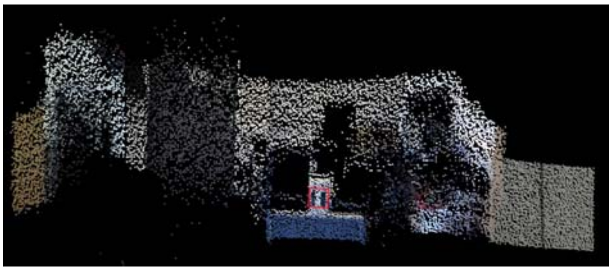


(a) (b)

Fig.13 – (a) MERLIN (b) output 2D/3D

The experiments test in the corridor inside the building. Figure 14 demonstrates the indoor 3D mapping in the corridor. The detail of fusion data between 2D/3D sensors have been enhanced in our previous work [13]. It can be seen that the mapping provides more texture information, which comprehends the user. Figure 14 (a-b) demonstrates the coordination between SLAM and the artificial landmark. Not only the system can generate the 3D mapping but also recognize the artificial landmark in the cutter environment. As these results, the mobile robotic can afterward apply to detect other objects, by training more interested objects causing the over all robot performance to be improved.

(a)



(b)

Fig.14 – Coordination of SLAM and artificial landmark recognition within 3D environment

8. CONCLUSIONS AND FUTURE WORKS

This paper presents the SLAM by using a new type of 3D sensor, PMD camera to cooperate with the artificial landmark recognition using Haar-like features in order to apply in various mobile robotic application tasks. The output illustrates that it is able to detect and to recognize the objects with different appearance, despite severe occlusions and cluttered backgrounds. The PMD camera is very attractive in terms of real time data capturing. To acquire more reliability of 3D texture, the fusion data between two sensors, 3D time of flight and 2D high resolution color camera are presented. The mobile robot locomotion is the correspondence position to yield the entire environmental data within tracking trajectory. This enables robot to control itself autonomously. However, the quality of output can be improved by using the better quality from 2D camera and by increasing the resolution of PMD camera that has been researched. The noise reduction will be presented in the next experiments. In the future work, the algorithms for the nearest rang reading (NR) or the nearest range reading considering color (NRC) will be proposed to reduce the noise in order to yield better and reliable output.

9. ACKNOWLEDGMENT

We would like to thank the German Academic Exchange Service (DAAD) for providing financial support in the form of scholarship grants.

10. REFERENCES

[1] R. Lange. 3D Time-Of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology. *PhD thesis, University of Siegen*, 2000.

[2] http://www.pmdtec.com

[3] Thorsten Ringbech, Bianca Hagebeuker. A 3D Time of Flight Camera for Object Detection. *Optical 3-D Measurement Techniques, 9-12July, 2007. Zurich.*

[4] Zhengyou Zhang. A Flexible New Technique for Camera Calibration. *Microsoft Research,*13 Aug,2008.

[5] Henry Medeiros, Hidekazu Iwaki and Johnny Park. Online Distributed Calibration of a Large Network of Wireless Cameras Using Dynamic Clustering, *Distributed Smart Cameras, 2008. ICDSC 2008*, 7-11 Sept. 2008. p. 1-10.

[6] Jean-Yves Bouguet, Ph.D. *http://vision.caltech.edu/bouguetj/calib_doc/index.htm #system.*

[7] Jianbo Shi, Carlo Tomasi. Good Features to Track. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR94) Seattle,* June 1994.

[8] Bruce D. Lucas. An Iterative Image Registration Technique with an Application to Stereo Vision. *Proceedings DARPA Image Understanding Workshop,* April 1981, pp. 121-130.

[9] A. Bruhn, J. Weickert and C. Schnorr. Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods, *Internatinal Journal of Computer Vision.* 211-231, 2005.

[10] Sara Mitri, Simone Frintrop,Kai Pervölz, Hartmut Surmann and Andreas Nüchter. Robust Object Detection at Regions of Interest with an Application in Ball Recognition, 2005.

[11] Paul Viola and Michael Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. *Conference on Computer Vision and Pattern Recognition,* 2001.

[12] Gaojin Wen, Zhaoqi Wang, Shihong Xia and Dengming Zhu. Least-squares fitting of multiple M-dimensional point sets. *The Visual Computer: International Journal of Computer Graphics,* Vol 22, Pages 387-398, June 2006.

[13] C. Joochim and H. Roth. Development of a 3D Mapping using 2D/3D Sensors for Mobile Robot Locomotion. *IEEE international conference on Technologies for Practical Robot Applications (TePRA),* November 10-11, 2008. Woburn MA, USA.