Abstract—In this paper, we present a 3D map building method based on stereo vision for autonomous land vehicle(ALV). We achieve the 3D coordinates under the camera coordinate system(C-CS) from the disparity map firstly. Then we develop the conversion between the C-CS and the ALV coordinate system(ALV-CS) to acquire the local 3D map which reflects the actual scene basically. And the method is highly real-time and meets the need of ALV navigation. Furthermore, combined with the binocular stereo vision system, the error of the method is analysed based on the 2D prediction error from camera calibration and the disparity error from stereo matching, and the error item that describes the depth error between the built data of the scene point and the actual one is computed. The experiment to compensate the built 3D map is done and the result shows that the item can improve the precision of 3D map effectively. Then combined with information from INS and GPS, a global 3D map that reflects the actual scene, is built based on local 3D map.

I. INTRODUCTION

RECENTLY, there has been a great interest in using autonomous land vehicle(ALV) to build 3D maps of the environment, and many studies about map building exist in the literature[1-4]. In many respects, ALV moves in an unknown environment and it will get the surrounding information to build a map which is used to do location and navigation. Map building is a precondition of the true autonomous navigation in essence.[4].

The ALV requires a real-time planner using the local map which radar is widely applied in because its high speed of detecting obstacles and frequency of updating data. But radar is active sensor and is easier to be found by the enemy in some special scene, such as battlefield. Besides, radar is seldom used in cross-country environment because of its lower angle resolution. In recent years, real-time stereo vision algorithms have been possible as the high performance processor appears and almost have the same detection performance with laser radar at a distance less than 100 feet. Furthermore, stereo vision system, which consists of more than two cameras, is passive sensor and has some advantages such as concealment because of no radiation to acquire the images, swift and accurate measurement, and non-scanning imaging which is propitious to the ALV moving in adverse terrain.

In this paper, we first describe the architecture of the used binocular stereo vision system in section 2, closely following the algorithm to build a local 3D map based on the stereo vision, we look at representing 3D scene and acceptable experimental result is given in section 3. In section 4, we analyse and compensate the error of the algorithm combined with the 2D prediction error and the disparity error in the stereo system effectively. In section 5, we build a global 3D map based on local 3D map algorithm combined with INS and GPS. Section 6 is conclusion.

II. STEREO VISION SYSTEM

In this paper, the 3D map-building method is discussed based on the binocular stereo vision system, as showed in Fig.1, in which the optical axes of the two cameras are parallel each other and the baseline is 525 mm. The size of the input stereo image pairs is 320×240 pixels. At such level of resolution, rectification works well enough that the residual vertical disparity is less than one pixel. The stereo image pairs are transformed into image pyramids by image sub-sampling and image similarity is measured by computing the sum-squared-difference(SSD) or 7×7 windows over a fixed disparity search range. Disparity is estimated by finding the SSD minimum independently for each pixel. Sub-pixel disparity estimate are obtained by fitting parabolas to the three SSD values surrounding the SSD minimum and taking the disparity estimate to be the minimum of the parabola. The resulting disparity map is smoothed with 3×3 low-pass

Huahua Chen is now with College of Communication Engineering, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, CHINA (e-mail: science_ant@hotmail.com). Zezhong Xu is now with College of Computer and Software, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, CHINA (e-mail: xuzz_1999@263.net).
filter to reduce noise from the sub-pixel estimation process. Small islands of bad matches occasionally can be contiguous with larger regions of good matches or completely disjoint from good regions. A simple blob coloring algorithm[5] can be used to reject small disjoint regions. Furthermore, it is possible to eliminate small regions of bad matches that are contiguous to larger, good regions by making the connectivity criterion in the blob coloring algorithm a threshold on the disparity gradient. The matching algorithm is run on the personal computer based on PIV CPU running at 2.4GHz under RedHat6.1 and its runtime is less than 100 ms. Fig.2 shows the right image of stereo pair and the disparity map.

III. LOCAL 3D MAP-BUILDING

That the disparity, which is the 3D scene information based on 2D image coordinates, should be converted to the 3D information in the ALV coordinate system(ALV-CS), on which the ALV moves dependent, requires not only the accurate scene information description from the map but also the map building algorithm in highly real-time.

A. Coordinates of the Object Point Conversion from the 2D Image Coordinate to the 3D Camera Coordinate

Fig.1 shows the binocular stereo vision system. C1 and C2 are the two cameras whose x axes are overlapping and y axes are parallel each other. I1 and I2 are image planes. o1 and o2 are the optical centres respectively. P1(P2) is the projection of the any object point P in the image plane I1(I2). P1 and P2 have the same y value after rectification. The difference between P1 and P2 in x axis is disparity d.

According to the pinhole model[6], the 3D coordinate of any object point P under one of the camera coordinate system (C-CS)can be written as:

\[
\begin{align*}
X_C &= b \times (U - U_0) / d \\
Y_C &= b \times (V - V_0) / d \\
Z_C &= bf / d
\end{align*}
\]

where \((U_0, V_0)\) is the image coordinate of the principle point and f is the focus length of the camera, and they can be acquired by calibration. The corresponding 3D world coordinate under the C-CS of each pixel in the disparity map can be achieved using (1).

B. Conversion between the C-CS and the ALV-CS

The each distance from the objects in the scene to the ALV and the each height of the objects need to be known in ALV navigation so that the ALV can decide to go on or detour. Besides working out the coordinates of the all object points in the scene, they should be converted into the ALV-CS.

Fig.3 shows the relationship between the C-CS \(O_cX_cY_cZ_c\) and the ALV-CS \(O_wX_wY_wZ_w\). Plane \(O_cY_cZ_c\) and plane \(O_wY_wZ_w\) can be parallel after the stereo vision system is fixed accurately. \(a\) is the pitch angle. From Fig.3, we can achieve Fig.4 which describes the conversion relationship between \(O_cX_cY_cZ_c\) and \(O_wX_wY_wZ_w\). In Fig.4, P is the any object point, and \(O_x\) is perpendicular to the paper outwards. \(O'\) is the projection of \(O_c\) and a coordinate system \(O'X'Y'Z'\), whose each axis is parallel to the each corresponding one of \(O_wX_wY_wZ_w\), can be constructed. \(O'Z'\) and \(OZ_c\) are coplanar, and \(O'X'\) is perpendicular to the paper outwards. Vector \(\vec{T}_{O'cO_c}\), which describes the translation between \(O'\) and \(O_c\), and \(\vec{T}_{O'o_c}\), which describes the translation between \(O'\) and and

\[
\begin{align*}
O_w \Rightarrow O'w \Rightarrow O'c \Rightarrow O_c
\end{align*}
\]
The corresponding local 3D map of the disparity map shown in Fig. 2(b). Actual data of the objects and experimental data are listed in TABLE I and II. Combined with Fig. 2(a), the built map reflects the actual scene basically. The runtime is less than 10 ms when the algorithm is run on the personal computer based on PIV CPU running at 2.4GHz under RedHat6.1.

**IV. ERROR ANALYSIS AND COMPENSATION**

Local 3D map building works out the points in 3D space using the model parameters from calibration and the disparity map. The reliability and the precision of the solution is affected by many factors such as 2D prediction error which includes the error of the pinhole model and image noise, and disparity error which includes foreshortening error, misalignment error and systematic error[7], for the process of the map building is a inverse transformation of projection and a uncertain problem. In this section, the error of the 3D map building algorithm will be discussed.

**A. Error Analysis of Binocular Stereo Vision**

The imaging model of binocular stereo vision is shown in Fig. 6. $O_1$ and $O_2$ whose distance is the length of baseline, are the optical centres of camera 1 and camera 2 respectively. $O_1Z$, $O_2Z'$ are the optical axes of the two cameras, and they are parallel each other. We suppose that the length of the baseline is $b$, and the focus length of the two cameras is $f$. In order to simplify computation, we only discuss one-dimensional image in which the point is two-dimensional. The process of 3D imaging is the extension of that of 2D one, and the discussion result is applicable too.

The coordinates of $O_1$ and $O_2$ are $O_1 (0,0)$ and $O_2 (b,0)$ respectively. Two image planes are at $Z=f$. The coordinates of two calibration points $P_1$ and $P_2$ are $P_1 (x_1, z)$, $P_2 (x_2, z)$. The undistorted projection positions of $P_1$ and $P_2$ on the image plane are $u_{11}$ and $u_{21}$ according to the pinhole model. The actual projection positions are $u'_{11}$ and $u'_{21}$ because of lens distortion. So the 2D prediction error $\varepsilon$ can be written as:

$$\varepsilon = \sqrt{(X_d - X_u)^2 + (Y_d - Y_u)^2},$$  \hspace{1cm} (4)

where $(X_u,Y_u)$, $(X_d,Y_d)$ are undistorted and distorted pixel coordinates.

The coordinates of new optical centres $O'_1$ and $O'_2$ can be obtained by solving the linear equation combined with $P_1$, $P_2$, and their corresponding 2D image positions $u'_{11}$, $u'_{21}$ and $u'_{12}$,$u'_{22}$. From the intersection of $O_1P_1$ and the image plane, $u_{11}(fx_1/z, f)$ can be obtained. Correspondingly, $u_{21}(fx_2/z + \varepsilon, f)$ and $u'_{21}(fx_2/z + \varepsilon, f)$ can be achieved. From the intersection of $u'_{11}P_1$ and $u'_{21}P_2$, we have $O'_1(z \varepsilon/f, 0)$. In the same way, $O'_2(b-z \varepsilon/(z-f),0)$. Suppose that the true coordinate of the point $P_3$ to be tested is $(x_3, z)$. The two projection positions of $P_3$ are $u'_{31}(fx_3/z + \varepsilon, f)$ and $u'_{32}(b-f(b-x_3)/z - \varepsilon, f)$. Considering the disparity error, the actual projection positions of $P_3$ are $u'_{31}$ and $u''_{32}$. Then the coordinate of $P'_3$ can be obtained.

Suppose that the disparity error is $\delta$ written as:

$$\delta = ae_f + me_m,$$  \hspace{1cm} (5)

where $a$ is non-zero disparity gradient and $m$ is vertical misalignment, $e_f$, $e_m$ are foreshortening error and misalignment error. And they can be written as:

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>COMPARISON IN DEPTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual depth(mm)</td>
<td>Experimental depth(mm)</td>
</tr>
<tr>
<td>7000</td>
<td>6749</td>
</tr>
<tr>
<td>9000</td>
<td>8554</td>
</tr>
<tr>
<td>12000</td>
<td>10977</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>COMPARISON IN HEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual height(mm)</td>
<td>Experimental height(mm)</td>
</tr>
<tr>
<td>450</td>
<td>516</td>
</tr>
<tr>
<td>510</td>
<td>622</td>
</tr>
<tr>
<td>810</td>
<td>960</td>
</tr>
</tbody>
</table>

**Fig. 5. Local 3D map for the disparity map shown in Fig. 2**
\[ e_f = \sum_{x=-3}^{3} \sum_{y=-3}^{3} \left( \frac{\partial I_x}{\partial x} \right)^2, \]
\[ e'_m = \sum_{x=-3}^{3} \sum_{y=-3}^{3} \left( \frac{\partial I_x}{\partial x} \right) \left( \frac{\partial I_y}{\partial y} \right), \]

So we have \( u''_{32} = (b-f(b-x_3)/z - e - \delta, f) \). And the length of the new baseline \( O'_1O'_2 \) is \( (b-2(z e - \delta)/z - e - \delta) \). We have new disparity \( d = f(b-x_3)/z + e + \delta + (fx_3/z + e)+2ze e - \delta) \). According to the relationship of two cameras in Fig. 6, the depth of \( P'_3 \) is

\[ z' = (b-2ze - \delta)/[f(b-x_3)/z + e + \delta + (fx_3/z + e)] - 2ze e - \delta - 2z^2 e - \delta \). Thus, the depth error of the binocular stereo vision under C-CS can be written as

\[ \Delta z = z' - z = \frac{z^2 \delta(f - z)}{fb(z - f) + z^2 \delta - 2z^2 e - \delta}. \]

Putting the depth error into (1), we will have new world coordinate under the ALV-CS correspondingly.

### B. Experimental Result

To show how (8) works, we simulate the scene shown in Fig. 1. Experimental result is listed in Table III and IV. Compared with TABLE I, II, III and IV, the result shows that (8) can improve the precision of 3D map.

#### TABLE III

<table>
<thead>
<tr>
<th>Actual depth(mm)</th>
<th>Compensated depth(mm)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>7000</td>
<td>6914</td>
<td>1.23%</td>
</tr>
<tr>
<td>9000</td>
<td>8795</td>
<td>2.28%</td>
</tr>
<tr>
<td>12000</td>
<td>11428</td>
<td>4.77%</td>
</tr>
</tbody>
</table>

#### TABLE IV

<table>
<thead>
<tr>
<th>Actual height(mm)</th>
<th>Compensated height(mm)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>450</td>
<td>486</td>
<td>8.00%</td>
</tr>
<tr>
<td>510</td>
<td>579</td>
<td>13.53%</td>
</tr>
<tr>
<td>810</td>
<td>892</td>
<td>10.12%</td>
</tr>
</tbody>
</table>

### V. GLOBAL 3D MAP BUILDING

When the ALV performs global planning, it must recur to a global map. A global 3D map has more information than a 2D one. We will build a global 3D map based on local 3D map combined with information from INS and GPS.

#### A. 3D Map Description with Grids

Data from above equations direct are unsuitable for describing global 3D map.

1. Data are mass and much of them are redundant. If there is a box in front of the ALV, the achieved data form a plane towards the cameras. The most important among those data to the ALV is that the representative data can describe the position and the height of the object. So we can make few data depict the scene.

2. Different space resolution of the camera imaging results in different data distribution in close regions and in far regions. In close regions 3D data is dense, and in far regions 3D data is sparse or even nonexistent.

3. Data loss results from occlusion. No data exists behind the obstacle that is higher than the floor and can not been seen by the camera.

In this paper, grids are introduced to describe the scene. The height of each grid, which is a region whose size is 30 cm x 30 cm, depends on that of data fall into the grid. The scene is quantified into regions, and the height of the position that data are nonexistent can be interpolated by other data in the same region. But the minimum depth error may increase to 30 cm even if only tiny depth error occurs, for example 5 cm.

#### B. Global 3D Map Building

Global 3D map must be in the same coordinate system, for example world coordinate system (W-CS). So when the current local 3D map is integrated into the global one, the ALV-CS must be converted to the W-CS. In this paper, W-CS is built based on INS, which determines the x axes and z axes of the W-CS, and GPS, which determines the
origin of the W-CS. Global 3D map can be built based on the local 3D map algorithm combined with information from INS, which provides pitch, roll and heading angle of ALV to correct the pose of ALV, and GPS, which gives the current position of ALV in the W-CS so that the conversion of scene information between the ALV-CS and the W-CS can be done.

We simulate the avenue shown in Fig.7. (a)-(f) are representative scene sequences. And built 3D map is shown in Fig.8. The result shows that built 3D map reflects the actual scene basically.

VI. CONCLUSION

In this paper, a local 3D map building algorithm for ALV is proposed based on binocular stereo vision and the algorithm is highly real-time and meets the need of ALV navigation. Then, a model for binocular stereo vision system is constructed, and combined with the binocular stereo vision system, the error of the algorithm is analysed based on the 2D prediction error from camera calibration and the disparity error from stereo matching. And the item for compensating the depth is provided. Simulation and experimental result is given to show that compensated error can improve the precision of 3D map in depth and height effectively. Based on the local 3D map building algorithm, a global 3D map is built combined with information from INS and GPS, furthermore its experimental result is given to show that the built global 3D map reflects the actual scene basically.

REFERENCES


