네트워크 상에서 비구성 환경의 원격제어를 위한 모델 기반의 스테레오 비전 시스템에 관한 연구

이 형 국*, 정 진 현
광운대학교 제어계측공학과

A study of a model based stereo vision system for a remote control in the unstructued environment on networks

Hyoung Guk Yi, Chin Hyun Chung
Dept. of Control and Instrumentation Eng., Kwangwoon Univ.

Abstract - To control the remote system in the unstructured environment requires data under certain circumstances. When a machine is dealt with an unstructured environment, new environment structure is to be composed. The stereo vision system can get both the intensity data and the range data. So, in this paper, data architecture of a stereo image is proposed to get them.

1. Introduction

A major topic in computer vision has concerned us to know how one can deduce information on the location and shape of objects in the three dimensional structure. Methods how disparity values are computed from two images during a process of the depth perception are referred as the binocular image-disparity stereo.

The communication system has caused a lot of changes of our working environment to get lots of information very rapidly through the networks. In fact, the development of network has enabled us to control the remote system. If there is the object we intend to control, we can obtain the location of the object in the unstructured environment through a stereo camera system.

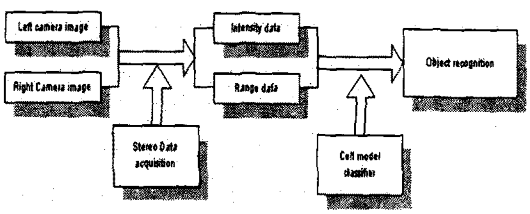

Fig 1. The structure of a stereo vision system

2. Stereo camera model

2.1 Geometrical camera model

Vision system is composed of two cameras placed with accurate focal-length in parallel base line. Stereo camera system acquires a depth information with disparity of two other projection image planes, so the coordinate of the cameras has to be defined. The real world coordinate of the object can be got from the world coordinate system of the camera coordinate system. Each camera creates its own coordinate system, but if one coordinate system is set as a reference coordinate, the other can be computed as new coordinate of a point that

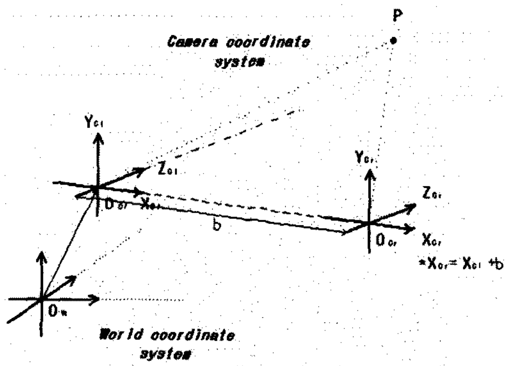

Fig 2 The camera and world coordinate system

can be rotated and translated. Reference camera coordinate is represented as following equations.

$$R_\alpha=\begin{bmatrix}1&0&0\\0&\cos\alpha&\sin\alpha\\0&-\sin\alpha&\cos\alpha\end{bmatrix}\quad R_\beta=\begin{bmatrix}\cos\beta&0&-\sin\beta\\0&1&0\\\sin\beta&0&\cos\beta\end{bmatrix}$$

$$R_\gamma=\begin{bmatrix}\cos\gamma&-\sin\gamma&0\\\sin\gamma&\cos\gamma&0\\0&0&1\end{bmatrix}\qquad(1)$$

$$P_{O_w}=R_\alpha R_\beta R_\gamma P_{O_{c_l}}+T_w\qquad(2)$$

Assuming the left camera as a reference camera, we can obtain the right camera coordinate system is

$$\begin{bmatrix}X_{C_r}\\Y_{C_r}\\Z_{C_r}\end{bmatrix}=\begin{bmatrix}X_{C_l}+b\\Y_{C_l}\\Z_{C_l}\end{bmatrix}.\qquad(3)$$

2.2 Depth

2.2.1 Detecting conjugate pairs using region correlation

Conjugate pairs are two points in different images that are the projections of the same point in the scene. Detecting the conjugate pairs in a stereo image is known as the correspondence problem. It can be stated as follows: for each point in the left image, find the corresponding point in the right image. Both edge features and region features are used in the stereo matching. But an important limitation of edge for stereo matching is that the value of the computed depth is not useful along the closed edges where the depth is not well defined. So one of the methods to find its potential feature is the region correlation method that identifies the interesting point and matches the conjugate pairs.

The correlation function has the disadvantage of being sensitive to changes in the amplitude of $f(x,y)$ and $w(x,y)$. For example, doubling all values of $f(x,y)$ doubles the value of correlation. The correlation function to overcome this difficulty is to perform matching via the correlation coefficient, which is defined as

$$\gamma(s,t)=\frac{\sum_x\sum_y [f(x,y)-\bar{f}(x,y)][w(x-s,y-t)-\bar{w}]}{\left\{\sum_x\sum_y [f(x,y)-\bar{f}(x,y)]^2\sum_x\sum_y [w(x-s,y-t)-\bar{w}]^2\right\}^{1/2}} \quad (4)$$

where $s = 0,1,2,\ldots,M-1$, $t = 0,1,2,\ldots,N-1$, $\bar{w}$ is the average value pixels in $w(x,y)$, $\bar{f}(x,y)$ is the average value of $f(x,y)$ in the region coincident with the current location of w, and the summations are taken over the coordinates common both f and w.

2.2.2 Calculation depth value

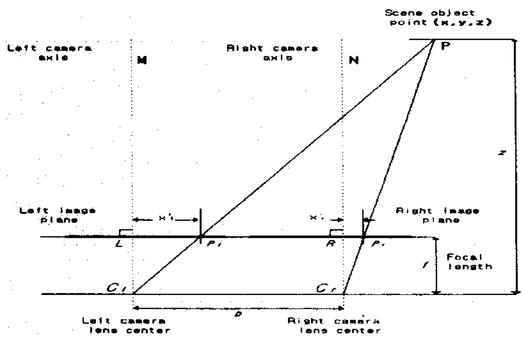

Fig 3 The calculus of the depth

In Fig. 3 the scene point P is observed at points p_l and p_r in the left and right image plane, respectively. Without loss of generality, we assume that the origin of the coordinate coincides with the left lens center.

Comparing the similar triangles PMC_l with p_lLC_l, we get

$$\frac{x}{z} = \frac{x'_l}{f} \quad (5)$$

Similarly, from the similar triangles PNC_r and p_rRC_r, we obtain

$$\frac{x-b}{z} = \frac{x'_r}{f} \quad (6)$$

Combining these two equations, we get

$$z = \frac{bf}{(x'_l - x'r)} \quad .(7)$$

and in the classic triangle determination we also obtain

$$z = \frac{\tan\theta_l \tan\theta_r}{\tan\theta_l \pm \tan\theta_r} \times b \quad (8)$$

where $\angle\theta_l = \angle PC_lC_r$, $\theta_r = \angle(180 - PC_rC_l)$

3. Classifier using the cell model

3.1 A general concept of a seed

The seed is defined as a unit of the object component. It is inserted at the feature region. Feature region is said to the place where intensity value changes rapidly or as interesting point as like edge, boundary, angle, corner and so forth. Seed is located in the central point, start point, endpoint and intersection point of the feature point. If feature contour is a simple shape or a closed loop, a simple seed can be located. If feature has inaccurate character and weak edge or high region-frequency, complex seed is spread. Therefor the class of the seed is variable. The square seed is mainly used. Its shape is quadrangle, and then handle easily with mathematics. The first square is inserted at each corner along boundary of the image plane.

3.2 Generation of the cell

Simply speaking, the position of each node is located at interesting point in the variable range. A choice of the range is important. If broad range data were got together, it is so hard to recognize the object because the image plane is simplified.

Between the two image regions, the general polynomial of the deformation function can be obtained as

$$d(p) = \sum_{0\le l,m\le N_1, l+m\le N_2} a_{l,m}x^l y^m \quad .(9)$$

where $p = [x,y]^T$ is the original image pixel coordinate point, $d(p) = [\xi(p), \eta(p)]^T$ is the deformed image pixel coordinate point.

In other words, if $N_1 = 1$, $N_2 = 2$ in the Eqn. (9) bilinear mapping polynomial is

$$d(p) = a_{00} + a_{10}x + a_{01}y + a_{11}xy \quad (10)$$

The mathematical description of the

germination of the square seed model is

$$\phi_1(u) = (1+x)(1+y)/4 \quad \phi_2(u) = (1-x)(1+y)/4$$
$$\phi_3(u) = (1-x)(1-y)/4 \quad \phi_4(u) = (1+x)(1-y)/4 \quad (11)$$

where $0 \le \phi_k(p) \le 1$, $\sum_k \phi_k = 1$.

The final cell polynomial description is

$$d(p) = \sum_{k=1}^{K} \phi_{K(p)} d_k. \quad (12)$$

The criteria for determining the reference node position of the seed is as follows:

- Reference node is located in the interesting point (isolate points, line, edge).
- Cell has the regulation: not seriously deforms and transmutates.

Generally features can be gotten by tracing the natural disposition of the edge and then we use sigmoidal nonlinear mapping to progress the natural disposition of the weak edge.

$$F(p) = \sigma\left(\left\|\frac{\partial\psi}{\partial p}\right\|^2\right) \quad (13)$$

$$\sigma(f) = \frac{1}{1+\exp(-\lambda(f-f_c))} \quad (14)$$

$$E_f = \frac{1}{2}\sum_{n\in N} F_n \quad (15)$$

As the number of frames increase, the original cell is compared with the next frame. If there is something transformed, the original cell is deformed by the parameter newly obtained and the deformed cell terminates the extinct cell.

The change of the cell can be determined by the find minimum matching error value between previous frame and current frame in the interframe. Let $\psi_i(p)$, $i = 1,2$ represent two frames of a video sequence. Assume that the displacements of the corner nodes from the second frame to the first frame are d_k. ,Eqn. (16) describes the displacement of an interior point $d(p)$ is related to node displacement. The MSE of the prediction for this block is described by

$$E_i = \frac{1}{2}\sum_{p\in D}(\psi_1(p+d(p)) - \psi_2(p))^2 \quad (16)$$

4. Experiment and result

In this paper, object data is made in the unstructured environment using a stereo vision system. Camera feature is specified that focal length is 0.26mm and the length of the base line is 15cm. The size of image plane is 400× 300. Stereo vision system can obtain the intensity data and the range data from two image planes with different viewpoints. The cell based classifier supplies less data size than data using statistical methods for translation data in the remote area, because of only translating the feature and vector of the cell. And Fig 6 and Fig 7 show a picture where the cell grows in the image plane. Fig 8 shows the translation data from the remote environment.

Fig 4. Original left image

Fig5. Original right image

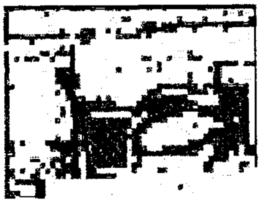

Fig 6. the Growing the cell in the left image

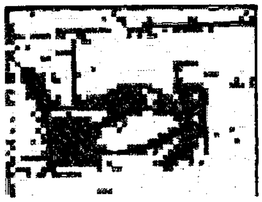

Fig 7. the Growing the cell in the right image

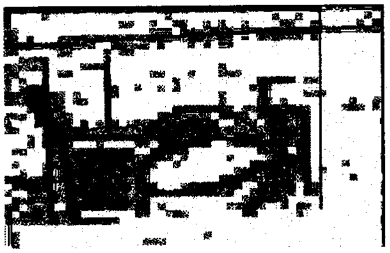

Fig 8. The Fusioned Image obtained by stereo vision system

[Reference]

[1] Antonio Francisco, "Continuous principal distance change for binocular depth perception", Image and Vision Computing Volume 13, 1995

[2] Jae-Woong Yi, Jun-Ho Oh, "Estimation of depth 3D motion parameter of moving object with multiple stereo images", Image and Vision Computing Volume 14, 1996

[3] Geoffrey S. Hubona, Gregory W. Shirah, David G. Fout "3D object recognition with motion" Rt96, 1996

[4] Paul Milgram, Anu Rastogi, Julius J "Telerobotic Control using Argumented Reality", IEEE, 1995

[5]Ouseb Lee and Yao Wang "Motion Compensated Prediction Using Nodal -Based Deformable Block Matching " IEEE 1995