IAPR Workshop on CV - Special Hardware and Industrial Applications OCT.12-14, 1988, Tokyo

# ACQUIRING RANGE DATA IN THE BINOCULAR 3–D COMPUTER VISION USING EDGE–BASED HIERARCHICAL MATCHING

Sheng Jin Yunming Li

Robotics and Artificial Intelligence Laboratory
China Textile University
1882 Yan–An Road(W), Shanghai, People's Republic of China

## Abstract

This paper deals with a binocular 3–D computer vision system based on the hierarchical matching of edge features, Frei and Chen operator is used to extract the edge. The average gradients of an image obtained by two isotropic operators are non–equal quantized and thresholded in an angle. Edge features are extracted after passing a preemphasis transfer function which can equalize the noise affection. Binary edge images are decomposed into a pyramid structure which is stored and searched using Illiffe's location method. Corresponding points are used to determine the range data using triangulation based on an improved Trivedi's formula. In calibration we set the optical axes of the two cameras parallel to simplify the calculation. A 3–rd order Householder transform is used to solve the compatible coupled equations.

## 1. Introduction

One of the main problem in 3–D computer vision is range measuring which has a broad sense. According to different usage, range measuring method could be divided into two categories. 2–D analysis approach and 3–D pattern approach. Triangulation passive is the method rather important and very similar to human vision in the 3–D pattern approach, it is also called binocular stereo vision.

The key point of the binocular vision is to determine the corresponding pixels of the same object point in a pair of images taken from slightly different view points, i.e. the matching problem. The matching methods can be divided into area–based and feature–based. As the former, Barnea and Silverman propose sequential registration[2]. Two–state template matching and coarse–fine template matching are suggested by Rosenfeld and Vanderbrug[3][4]. Wong and Hall extend the coarse and fine template matching to hierarchical matching[5]. Area–based algorithm is simpler but more sensitive to geometric distortion caused by binocular stereo disparity has stronger tendency to have mismatching and error–matching and higher computation cost. The feature–based registration includes statistical, edge and structure approach. The frequency domain features are used in the statistical feature approach, edge approach utilizes the distinct parts of the gradient change as feature, the result come from 2–D pattern recognition, such as nodes, links, circuits and the relations among them constitutes the features of structure approach. The feature–based registration is faster and not sensitive to the photometric and geometric distortion, as well as the parameter disparity of the video cameras. The problem is neither the poles and link segments nor the star structure and circuits can efficiently describe the 3–D objects and is the high computation cost.

This paper deals with an edge–based hierarchical matching approach. Frei and Chen operator is adapted, the angle between an image vector and its projection onto the edge subspace is used as the criteria. Non–equal quantization is applied to the convolution result of the edge operator to strengthen the weak edges. preemphasis transfer function is used to map the gradient vector into the edge subspace to raise the edges in the high gray level location but keep depressing noises in the low gray level location. Complete quad tree with Illiffe's location method is used for image compression and speeding up the addressing speeds. Hierarchical matching starts finding corresponding points at the lowest resolution level, then searches at the next higher resolution level until the highest resolution is reached. The improved Trivedi's formula is used to determine the range data. A 3–rd order Householder transform is used to solve the compatible coupled equations.

## 2. Edge extraction, non–equal quantization and preemphasis

In the image space, edge extraction uses a lot of weighting function to convolute and gets result edges when the gradient exceeds a threshold. Frei and Chen propose a set of $3 \times 3$ complete orthogonal basis, get the edge as the angle between the image vector and the edge subspace is greater than a threshold, i.e. judge the pixel as edge if

$$\Theta = arc\ cos\left[\sum_{i=1}^{e}(B,W_i)^2 / (B,B)\right]^{1/2} \quad < \Theta_t,$$

B is the image vector, $W_i$ the ith operator, e the dimension of the edge subspace. Convolution is very time–consuming, consider the contribution of each operator to the subspace, we give the algorithm for the edge extraction.

1) Correlate template B with two isotropic average gradient operators $W_1$, $W_2$: $\sum_{i=1}^{2}(B,W_i)^2 / (B,B)$.

2)Repeat 1), convolute the original image, get the gray scale edge image.

3)Convolute the gray scale edge image from 2) with point template, set two thresholds $T_1$ and $T_2$, if $T_1 <$ convolution output $< T_2$, the pixel where the template locates is edge.

The above algorithm could save the convolution cost of 6 operators, the key points are 1. $W_1$, $W_2$ give the extrema of the gradient, 2. edges are non–isolate extrema points of the gradient. The angle

$$\Theta = arc\ cos\left[\sum_{i=1}^{e}(B,W_i)^2 / (B,B)\right]^{1/2} \quad \text{is}$$

normally equal–quantized, the $q_1$ in Fig. 1 is the quantization curve. The entropy of such coding is small, the method to increase entropy is to get the optimal coding according to the different angle value. the image gradient $(B,W_i)$ depends on the contrast and gray level of the image which can be got from histogram analysis. Set one or two small probability value, take an interval between the maximum and minimum gray level, estimate the gradient distribution by the aid of extrema analysis of the histrogram, nonlinear quantization could be done using these information. A simple method is to equal–quantize in the interval between maximum and minimum gray level and set zero to the gray level which is impossibly to appear, the quantizing curve

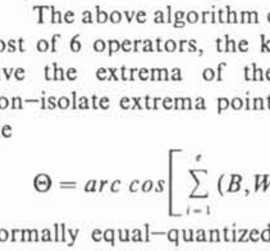


**Fig.1 Three quantization curves.**

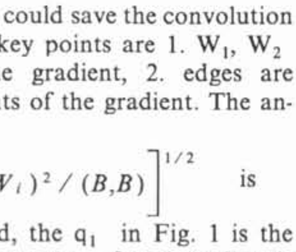


**Fig.2 The principle of preemphasis.**

is $q_2$. The other better quantization method chooses broken lines with different slopes in the interval between maximum and minimum gray level, as $q_3$ in Fig. 1. The advantage of non–equal quantization is its uniformly distributed gradient histogram, i.e. the gradient is well distributed, the edges with small gradient value can be extracted clearly.

Using angle as criteria, a new problem, the influence of the noises which are mixed during inputting and transmitting the image appears. The noises cause some extrema of the gradient. The introducing of the (B,B) in the denominator strengthens the edges with small gradient value, but causes the gradient very sensitive to the noises. According to the matched filtering theory, in the case of white noise the matched–filter with the highest signal to noise ratio(SNR) is the mirror function of the original function. In the case of colour noise as $N_\Theta(B)$ in Fig. 2 in order to get the highest SNR, a preemphasis transfer function $H_1(B)$ is designed (see Fig.2) to get the $H=H_1H_2$ a highest SNR output. Here, $H_2$ is the transfer function of the matched filter in the case of white noise. It is difficult to design a preemphasis transfer function of which high order derivative is continuous. In practice, approximation with segment by segment broken lines is taken which gets satisfied result. The computation cost which is addition level is rather low.The advantage of the preemphasis is to raise weak edges and to depress the noise influences at the same time.

### 3. Hierarchical matching, sequential registration and threshold analysis

Hierarchical scene matching utilizes the correlation of the correlation function. Coarsely search the corresponding point in the lowest resolution level, then accurately match in the next higher resolution level only in the location close to the corresponding point. A lot of non–corresponding points could be excluded in the lower level thus saves the computation cost. Hierarchical matching based on the quad tree of the image is computationally more efficient. Suppose the image has pixels $N \times N$ with a template sized $M \times M$, there is $(N-M+1)^2M^2$ computation cost for the point to point correlation. If we reduce the image resolution, the most possibly reduced level is $\log_2(N-M+1)$. The computation cost for every pixel in a window is 4 times as large as that of the order of the level. Consider the position combination, the computation cost is divided by 2, the cost for every pixel is $\log_2(N-M+1)^2$. As the matching window is $M \times M$, the total matching computation cost is

$K_1M^2\log_2(N-M+1)^2$, here $K_1$ approximates to the average number of the corresponding points in the lowest resolution. The reduction computation cost is $2\sum_{i=1}^{L-1}\left(\frac{N}{2^i}\right)^2$, then the computation efficiency is:

$$\eta = \frac{(N-M+1)^2M^2}{\left[K_1M^2log_2(N-M+1)^2 + 2\sum_{i=1}^{L-1}(\frac{N}{2^i})^2\right]}$$

$$= \frac{(N-M+1)^2}{\left[2K_1log_2(N-M+1)^2 + 2\sum_{i=1}^{L-1}\frac{1}{M^2}(\frac{N}{2^i})^2\right]}$$

**Table I Experimental results.**

| No. | x mm | y mm | z mm | $R_x$ mm | $R_y$ mm | $R_z$ mm | Time Sec. | № of corresponding points in each level | illumination |
|---|---|---|---|---|---|---|---|---|---|
| 1 | −8 | 17 | 1412 | −7.1 | 14.9 | 1414.8 | 1.76 | 5,3,3,1 | ① |
| 2 | −14 | −12 | 1535 | −14.2 | −12.5 | 1538.6 | 1.38 | 5,5,2,1 | ① |
| 3 | 22 | 11 | 1412 | 23.8 | 8.5 | 1407.6 | 2.26 | 6,5,3,1 | ② |
| 4 | 13 | −14 | 1535 | 16.2 | −16.1 | 1528.9 | 2.63 | 7,6,5,1 | ② |
| 5 | −8 | −47 | 1500 | −6.9 | −45.6 | 1504.2 | 1.87 | 6,5,3,1 | ③ |
| 6 | 9 | −51 | 1500 | 7.7 | −54.1 | 1495.1 | 1.75 | 6,5,5,1 | ③ |

① 4 fluorescent lamps + natural light.

② 2 fluorescent lamps + natural light with curtain.

③ 2 fluorescent lamps + natural light.

Experiments show usually $K_1 = 3 \sim 10$, then the hierarchical matching efficiency is $500 \sim 1000$.

The algorithm of hierarchical matching is:

1)In the search level L, using matching template $\frac{M}{2^L}$ correlate pixel to pixel in the location of $\left(\frac{N}{2^L} - \frac{M}{2^L} + 1\right)^2$, if the correlaton value is over a threshold, record that pixel as a corresponding point in this level.

2)Fine match every pixel in the location close to the corresponding point in the next higher resolution level, if the correlation value is greater than a threshold in this level, record that pixel corresponding point in this level.

3)Repeat 2), stop until any of the followings are met:

a)The 0 level(original image) is reached. The pixel with the largest correlation value is taken as the matching point.

b)If there is only a single corresponding point, that point is the matching point.

c)If there is no corresponding point in any level, there is no matching point at all. System fail to match.

Another computation efficient technique is sequential registration.[2] There are two kinds of threshold, constant value threshold $T_0$ and monotonic increasing threshold T(n). It is obvious the latter is more efficient in saving computation. A special $T^L(n)$ have to be chosen for every search level L, see Fig. 3. The decision of $T^L(n)$ is influenced by noise, probability of mismatching, probability of error–matching. In the wider sense, noises influence image feature, therefore it does not appear in Fig. 3. $\tau$ is the ratio of number of feature pixels to number of the whole pixels in a window. In the information abundant region, such as regions with several detail edges, the mismatching probability and the error–matching probability is lower, a smaller $T^L(n)$ is chosen. The threshold is determined by maximum likelihood ratio, the lost function is constant. If mismatching is not allowed in practice, increase the lost function, the mismatching probability curve is raised, the cross point moves to right, then a greater threshold has to be chosen.

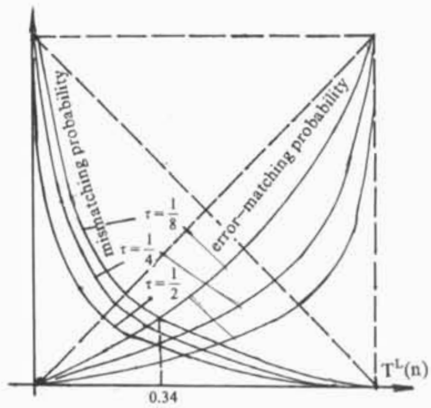


**Fig. 3 The decision of the threshold $T^L(n)$.**

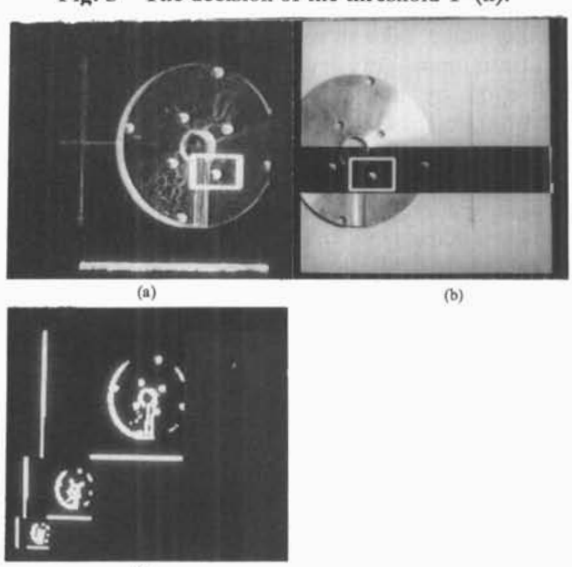


**Fig. 4 The matching process of a point.**

## 4. Result analysis

The matching process of a point is shown in Fig. 4, (a) and (b) are the left and right view point gray scale edge image, respectively. The rectangle frame in the left image is the chosen template, the rectangle frame in the right image is the matching result. (c) is the image reduced to 128 × 128, 64 × 64 and 32 × 32,respectively, from the binary image of (a).

Table Ⅰ gives the experimental results of 6 points. Table Ⅱ is the statistical analysis of the results.

**Table Ⅱ The statistical analysis of the results.**

| | x mm | y mm | z mm | Time Sec. | 4th level | 3rd level | 2nd level |
|---|---|---|---|---|---|---|---|
| Mean | 0.92 | −1.48 | −0.77 | 1.94 | 5.83 | 4.83 | 3.5 |
| Variance | 1.43 | 1.51 | 4.35 | 0.40 | 0.69 | 0.90 | 1.12 |
| Maximum | 3.20 | 3.10 | 6.10 | 2.63 | 7.0 | 6.0 | 5.0 |

From Table Ⅱ we can find:

1) The mean of the error shows the system calibration error.

2) The variance of the error shows the measuring error.

3) The error in the z axis direction is 2.96 times as large as that in the x and y axis direction. In order to improve the measuring accuracy in the z direction, the distance between two cameras have to be increased, which increases the disparity of the two images, causes the matching to be more difficult.

4) The mean of time shows the average reaction time of the system.

5) The average numbers of the corresponding points in each level show the threshold $T^L(n)$ excludes most of the search points, the increase of the varience of each level shows after fine matching the deletion result gets more obvious.

6) The object feature is rich when the illumination is abundant, which comes with more deletion points and faster matching.

## 5. Summary

We constitute a 3−D computer vision system on IBM PC / XT to get range data. The accuracy in x and y direction is 1.47mm and 4.35mm in z direction. The system reaction time is 1.94s. Triangulation passive approach faces two main problems, matching time and accuracy. Raising the image resolution and increasing the distance of the two cameras increases the accuracy, but the former increases the matching time obviously and the latter increases the difficulty of matching. It is shown in some extent the demand of matching time and the demand of accuracy is contrary to each other. To solve these problems, parallel processing and new features, such as star structure and circuits have to be considered.

## References

[1] S. Jin and Y. Li, "Range acquisition in 3−D computer vision," *Robot*, No. 6, 1988.(In press)

[2] D. I. Barnea and H. F. Silverman, "A class of algorithm for fast digital image registration," *IEEE Trans. Comput.*, Vol. C−21, pp. 179−186, 1972.

[3] G. J. Vanderbrug and A. Rosenfeld, "Two−stage template matching," *IEEE Trans. Comput.*, Vol. C−26, pp. 384−393, 1977.

[4] A. Rosenfeld and G. J. Vanderbrug, "Coarse−fine template matching," *IEEE Trans. Syst., Man, Cybern.*, Vol. SMC−7, pp. 104−107, 1977.

[5] R. Y. Wong and E. L. Hall, "Sequential hierarchical scene matching," *IEEE Trans. Comput.*, Vol. C−27, pp. 359−366, 1978.

[6] S. Jin and Y. Li, "Feature−based hierarchical registration of 3−D vision using star structure and circuits," *9th ICPR*, 1988.(to be presented)

[7] H. P. Trivedi, "A computational theory of stereo vision," *Proc. ofCVPR*, pp277−282, 1985.

[8] B. Zhang, *Linear algebra and its application*, Nanjing University press, pp308−310.

[9] W. Frei and C. Chen, "Fast boundary detection: A generation and a new algorithm," *IEEE Trans. Comput.*, Vol. C−26, pp. 988−998, 1977.