A Region-Based Stereo Algorithm

Gang Xu*f Hideki Kondo and Saburo Tsuji

Department of Control Engineering, Osaka University
Toyonaka, Osaka 560, japan
*also AI Department, ATR Communication Systems Research Laboratories
Seika-cho, Soraku-gun, Kyoto 619-02, Japan

Abstract

Feature- point- based stereo and segment-based stereo have been intensively studied. But why not region-based stereo? Region is more characteristic than point and segment. Matching between regions is easier and faster, if region segmentation is successful. Unfortunately, however, it is the failure to satisfy this precondition that has been the obstacle. Region segmentation in practice is subtle and unstable, and it is not guaranteed that region boundaries in stereo images correspond to same physical positions. Our approach to tackling this problem is to sharpen intensity contrast between regions by projecting a random pattern to the visual field. The proposed algorithm has been implemented to match both computer-generated random pattern stereo images and pattern-projected real images. Results show that the algorithm is robust, fast and tolerant of small vertical disparities.

1. Introduction

Stereo has been a hot topic in computer vision [Xu et al., 1987]. Feature-based stereo is currently the mainstream [Grimson, 1981a; Medioni and Nevatia, 1985; Ohta and Kanade, 1985]. The central problem is the correspondence problem, or the "false targets" problem; that is, for a feature in one image there are two or more match candidates in the other [Marr & Poggio, 1979; Crimson, 1981a,b]. At the beginning, feature meant point; points are matched between the images [Grimson 1981a,b], Later, segment-based stereo — segments rather than individual points are matched — were also intensively studied [Grimson 1985; Medioni and Nevatia, 1985; Ohta and Kanade, 1985]. Matching becomes easier because segment is more characteristic than point. A spontaneous extension, then, was to try region-based stereo; that is, regions are matched between images (it was first suggested in [Barrow &i Popplestone, 1971].) Region is even more characteristic than segment. Each region has a long vector of descriptions, such as area, perimeter, centroid, and so on. Matching based on these descriptions would have a much less probability of false targets provided that the images are not full of repetitive patterns. Anyway, it appears that everything is ready.

The only thing that we have overlooked above is region segmentation. In fact, it is the failure to satisfy this precondition that has been the obstacle. Region segmentation in practice is subtle and unstable, and it is not guaranteed that region boundaries in left and right images correspond to same physical positions. Thus, to fully appreciate the ease of matching between regions, we have first to somehow overcome the difficulty of segmenting stereo images into regions that strictly correspond to same physical positions.

Our idea is to sharpen intensity contrast between regions by projecting a random pattern to the visual field. The reason for preferring a random pattern is that it makes each region statistically distinguishable from its neighbor regions, while repetitive patterns suffer the "false targets" problem [Echigo and Yachida, 1985; Sato and Inokuchi, 1987]. We share the idea of random pattern with [Nishihara, 1983], but the two approaches have fundamental differences: we match regions that are projected from the pattern; Nishihara calculates correlation values between sign images (that are obtained by filtering the original image with a Gaussian and then binariziug it,) and does not rely on features in the projection pattern. The algorithm is particularly suitable for acquiring range data of industrial parts, because a projected random pattern can also make a dense texture on the surfaces, which industrial parts usually do not have themselves. The matching algorithm is fairly general and not restricted to pattern-projected images. Any stereo images that can be segmented to satisfy the special requirement can be matched by this algorithm, although it seems that there are few such images.

2. The Random Pattern Images

The algorithm consists of four steps:

(l)segment images into regions;

(2)calculate area, perimeter, centroid, y-limits, and number of holes for each region;

(3)match regions in the left and right images; and

(4)assign disparities to vertical boundary points.

Before describing the four steps in detail, we first introduce the computer-generated random pattern to be projected onto the scene. A random pattern image is obtained by filtering a computer-generated random dot image with a Gaussian of a suitable size and then thresholding it. An example is shown in Fig. 1. Compared with a random dot image, it has connected dot clusters of limited size and random shape, which are desirable properties for region-

f The work was done at Osaka University. Dr. Gang Xu's address from this autumn is Center of Information Science, Peking University, Beijing, China

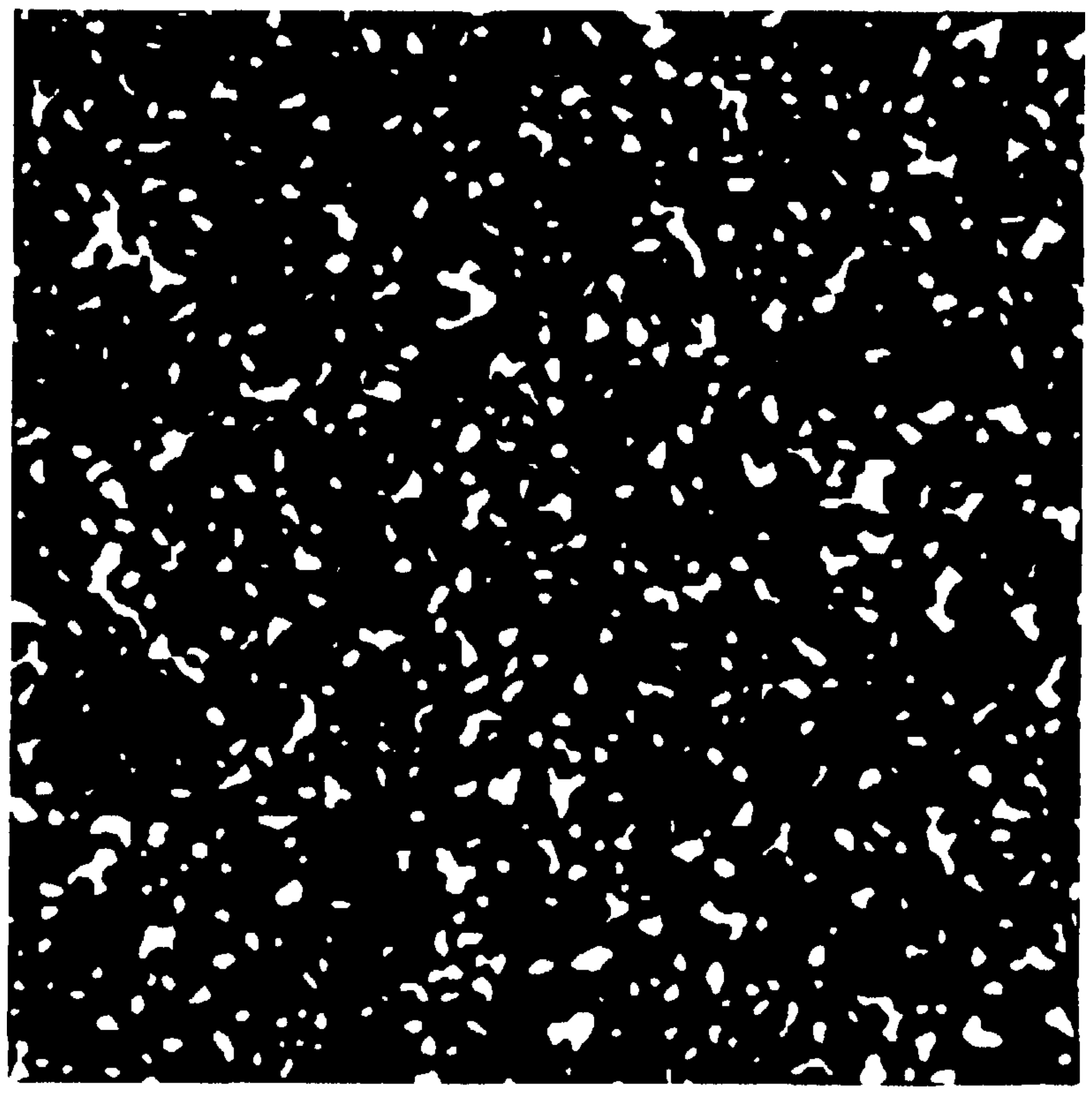


Fig. 1 The random pattern image

based matching. One problem is the size of regions in the pattern. A larger region will be more statistically significant and thus makes matching easier, but is more likely to cross an occluding boundary, and makes matching more difficult. There is a trade-off. To lessen the difficulty of matching the regions that cross an occluding boundary, we regulate the size of the Gaussian filter and the threshold to restrict the size of regions as far as it does not pose any problem to the disambiguation power of the matching process. Also the size of regions in the images can be changed by adjusting the focal length of the projector.

3. Region Segmentation and Region Descriptions

The stereo images are taken in low ambient light. We assume that the albedo of objects in a scene is uniform. Because of the sharp intensity contrast, simple thresholding is sufficient to separate bright pixels from dark ones. If the ambient light cannot be kept low enough, then we take four images, two before and two after the projection. Subtracting the two images on each viewpoint precedes thresholding and makes it easier. The next step is to define and number bright regions by tracing boundary pixels. Only the bright regions are matched between the left and right images, while the dark pixels are regarded as a dark background. Disparities along bright boundaries and disparities along dark boundaries are identical. The 4-connectedness is used to make it easy to separate different regions. Boundary pixels are distinguished from inner pixels. The image edges are treated as boundary pixels when needed. Being able to define each region perfectly, we can avoid the troubles of allowing one region in one image to be matched against, say, two regions in the other, and vice versa.

Once the images are segmented into regions, a vector of descriptions is calculated for each region [Ballard & Brown, 1982]. Area is simply the sum of the pixels of a region, and perimeter is the sum of the pixels of that region's boundary. Centroid is calculated by the following equations:

$$xc = \frac{\Sigma x}{area},$$
$$yc = \frac{\Sigma y}{area}. \qquad (1)$$

The y-coordinates of the highest and lowest points, referred to as yh and yl, respectively, are also recorded to indicate the region's height. In case where a hole exists inside the region (usually no,) it is also recorded.

4. Region Matching

The matching algorithm assumes, as the other stereo matchers often do, that the two cameras are so set that the epipolar lines are horizontal. But in practice it is only approximately satisfied. Thus it is desirable for the matcliing algorithm to be tolerant of small vertical disparities. Because of region's relatively large area, region-based matching deals well with small vertical disparities, and it is one of the advantages of matcliing larger entities, including both regions and segments.

Matching proceeds from the No. 1 region in the left image. For a region whose centroid is at *(xc, yc),* the search range in the right image is a rectangle with the length a little longer than the designed maximal disparity, as shown in Fig. 2. Any regions in the search range whose centroids are in this rectangle are regarded as candidates. The true match out of the candidates must simultaneously satisfy the following conditions:

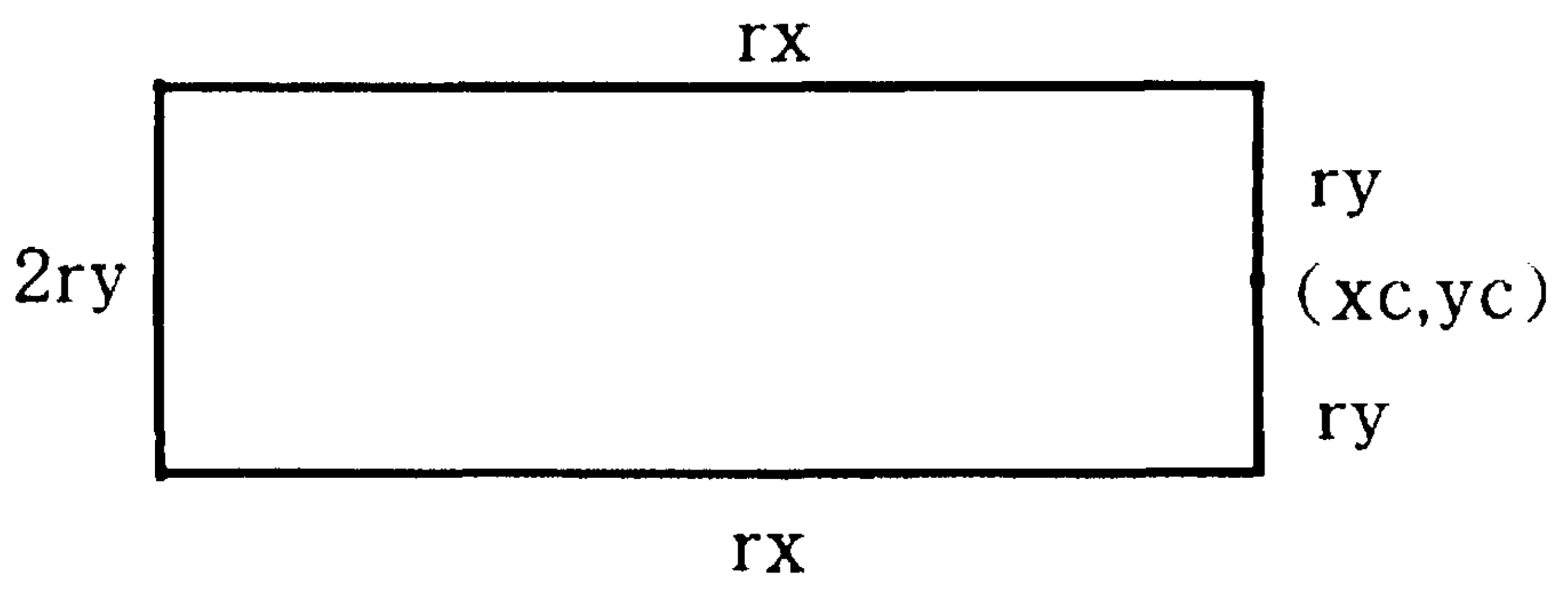


Fig. 2 A rectangle search range

$$\frac{|area_l - area_r|}{\min(area_r, area_l)} \leq T1;$$
$$\frac{|perimeter_l - perimeter_r|}{\min(perimeter_l, perimeter_r)} \leq T2;$$
$$|(yh_l - yl_l) - (yh_r - yl_r)| \leq T3;$$

The reason for these inequalities is that area, perimeter and height are the relatively invariant parameters of regions in the two images. The differences depend mainly on the viewing angles of the two cameras (see Appendix for a discussion.) The thresholds $T1$, $T2$ and $T3$ is to be determined in Section 6. If there are two candidates satisfying the inequalties, then we go to the next region to see whether one of the two candidates is this region's match. If so, then the remaining candidate is regarded as the unique match of the previous region. This has been called the "pulling effect" [Grimson, 1981a,b].

5. Disparity for Boundary Points ana Dealing with Occlusion

Once two regions are matched, disparity values are measured along the boundary. Since disparities for horizontal boundary points is meaningless, they are neglected. By horizontal boundary points we mean those points whose two closest neighbors are on the same horizontal line. (Recall that 4-connectness is employed to trace boundaries.) The disparity between the two centroids is used as an estimate to align the two boundaries. Inspite of a possible small vertical disparity, we still draw completely horizontal epipolar lines. The non-horizontal boundary points are matched by the orders that they appear on the horizontal epipolar lines. A histogram is drawn to show the distribution of disparity values along a boundary. If the region does not cover two depths, then disparity cannot change drastically along a single boundary (this is an application of the *continuity constraint.)* Only the points with small deviations (say, no more than 2 pixels) from the peak value are assigned disparities, while those with large deviations are discarded as noise.

Along depth discontinuities, two regions in the left image may merge into one in the right image, or reversely, one region in the left image may split into two in the right image. These regions basically cannot be matched by the above procedure, but are easily distinguished from the matched ones. If the limits in y-direction (max(yh1, yh2), min(y/l, *yl2))* and the sum of the areas of the two regions in one image are nearly the same as those of the region in the other image, then the one-to-two region correspondence relations are determined instead of one-to-one relations. Again completely horizontal epipolar lines are drawn. As shown in Fig. 3, if an epipolar line intersects two regions, then the two ends are selected as the matches of the boundary points on the same epipolar line in the other image. The selection is correct as far as the disparity is not large enough to reverse the orders that any points appear in the two images. A histogram is drawn to show the distribution of disparity values along the boundary. There should be two peaks, because the region crosses two surfaces in different depths. Points far away from either of the two peaks are discarded as noise.

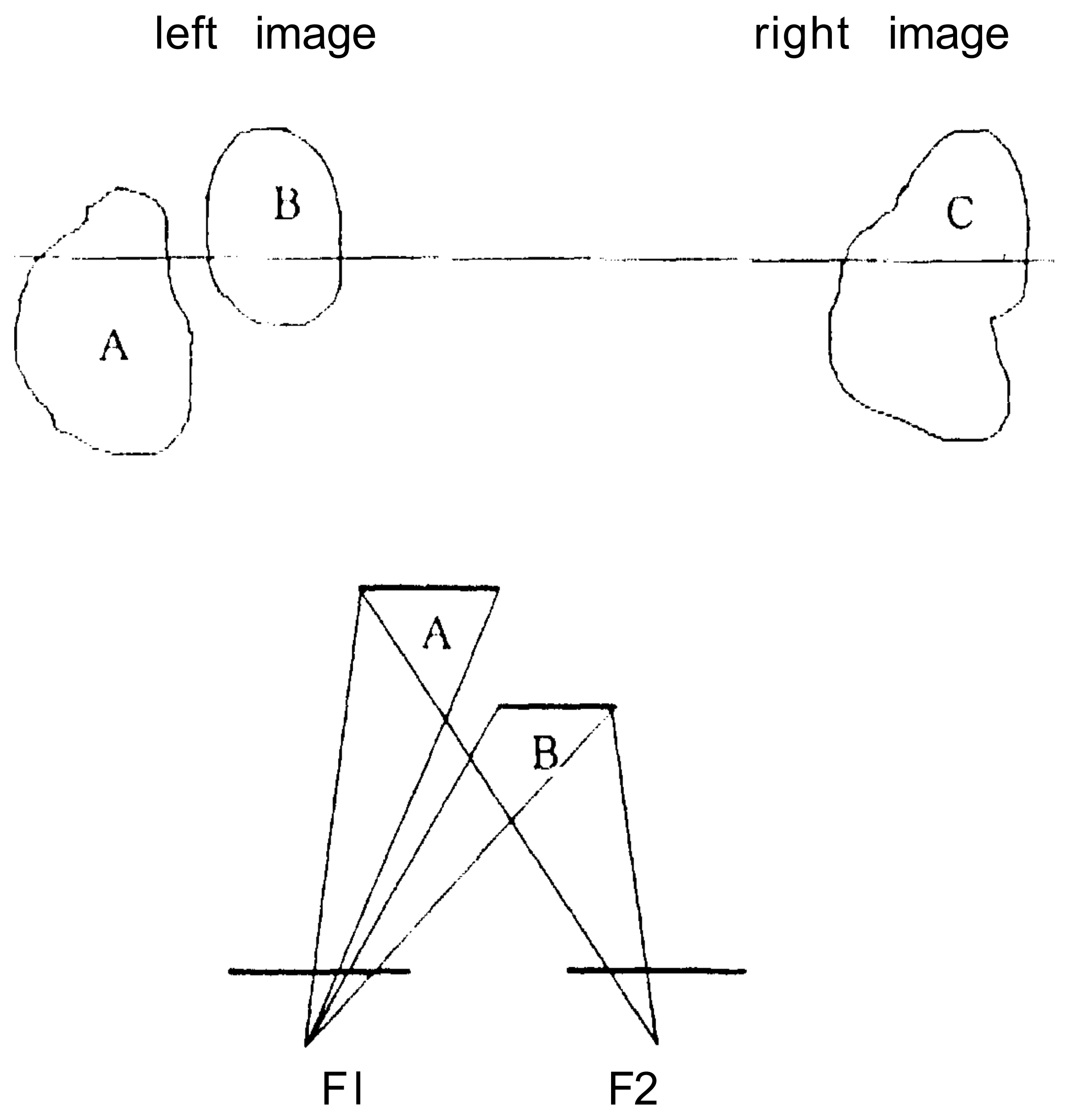


Fig. 3 Contour matching for regions crossing an occluding boundary

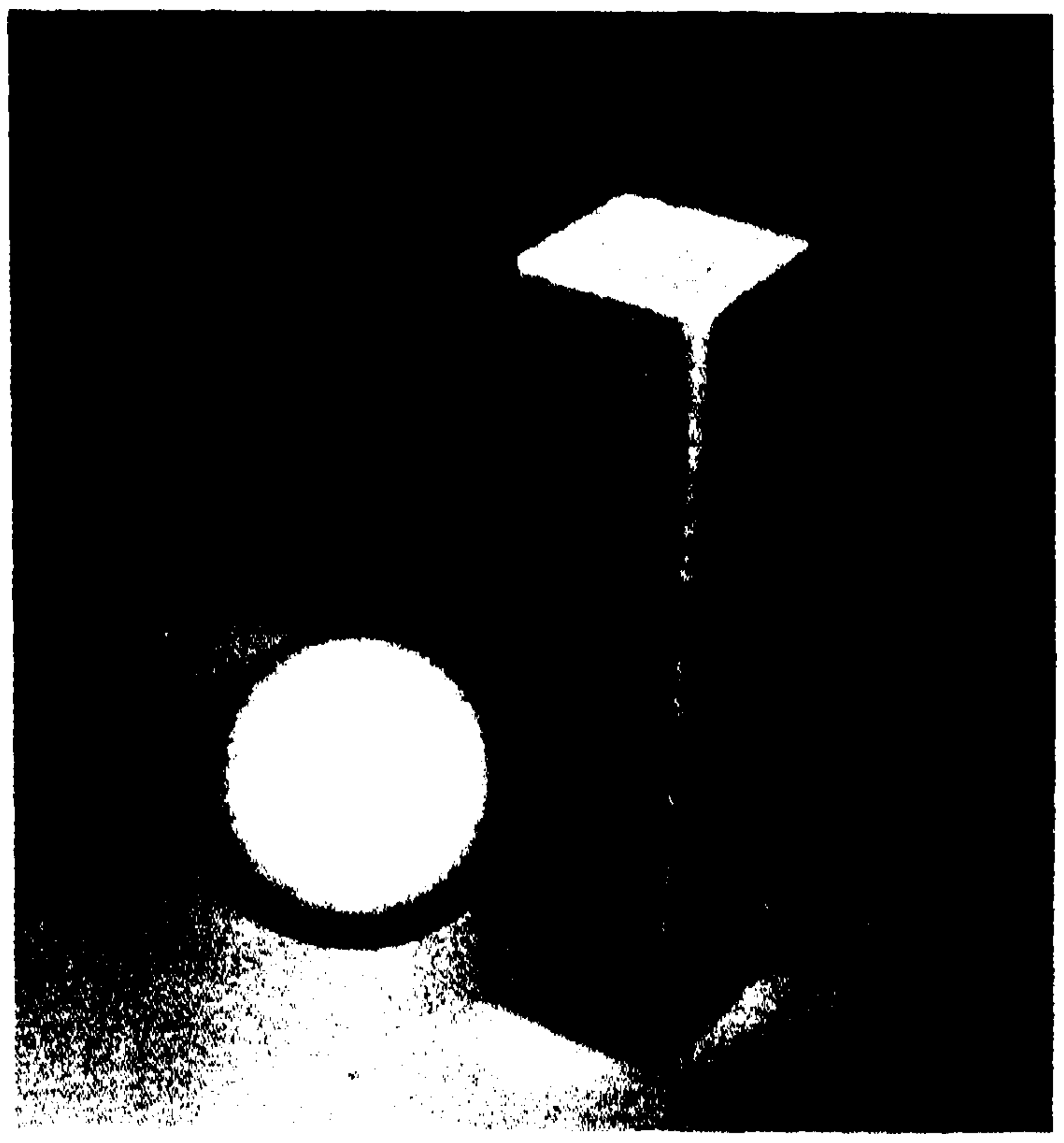
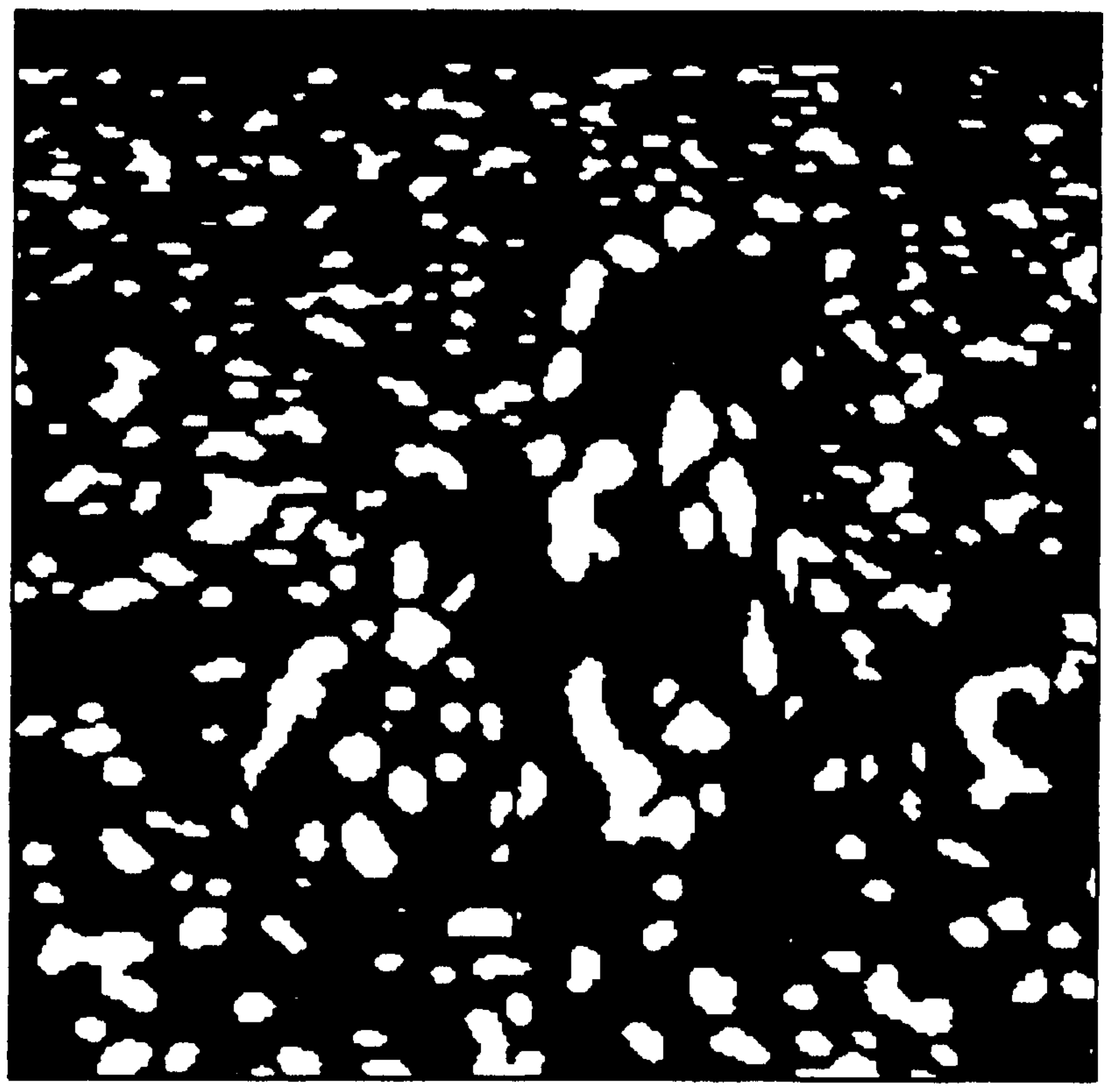


Fig. 4 The scene as the pattern is not projected

6. Implementation and Results

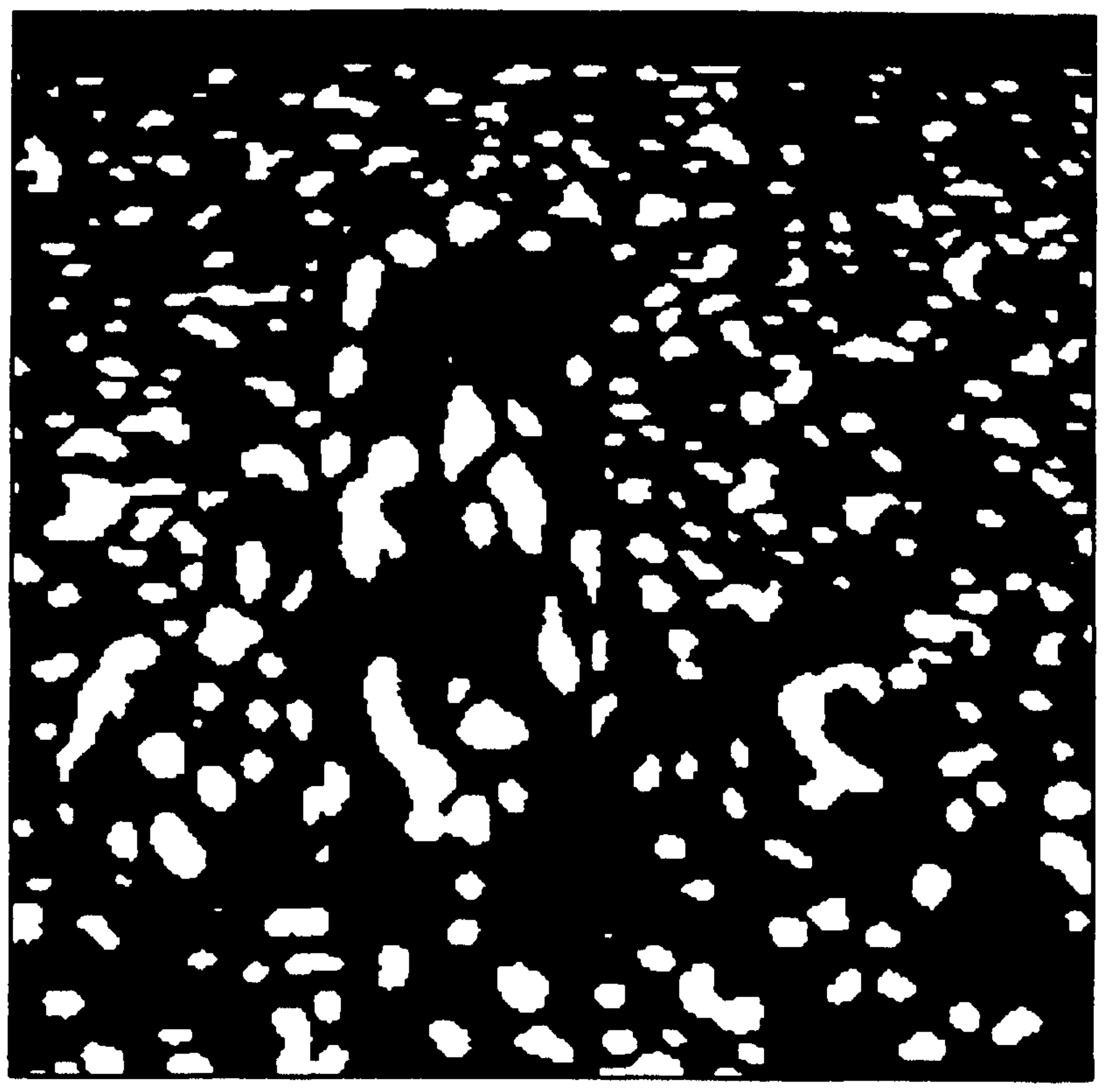
The algorithm has been implemented on a Sun-3 workstation and a Tospix-2 image processor. Thresholding, numbering, attribute calculation are executed by the image processor, and region matching and contour matching by the workstation. Both computer-generated random pattern stereo images and pattern projected real stereo images have been successfully matched.

An image taken before the pattern projection is shown in Fig. 4. The image shown in Fig. 1 is used as the random pattern and projected onto the scene. The stereo images are taken as the pattern is projected. The image size is 256 by 256 pixels. Their thresholded images are shown in Fig. 5. There are 291 and 300 regions in the left and right images, respectively. The image processor spends less than 2 sec. and the workstation spends about 12 sec. (region matching 2 sec. and contour matching 10 sec). The length and width of the search range are 65 and 3 pixels, respectively. There are in the left image 143 regions that each have only 1 candidate, 82 regions that each have 2 candidates, 16 regions that each have 3 and more candidates, and 46 regions that have no candidate, in their respective search ranges. After applying the inequalities and the "pulling effect", the situation changes to: 216 regions have only one candidate (the match), no regions have 2 or more candidates, and 75 regions (most of them are along the left edge of the image) have no candidate, in their respective search ranges. In the equalities, the thresholds for area, perimeter and height are 0.5, 0.5 and 2 pixels, respectively. Disparity values are measured along matched boundaries. A histogram of disparity values is taken for each bound-



(a) left

Fig. 5 The thresholded stereo images



(b) right

Fig. 5 The thresholded stereo images

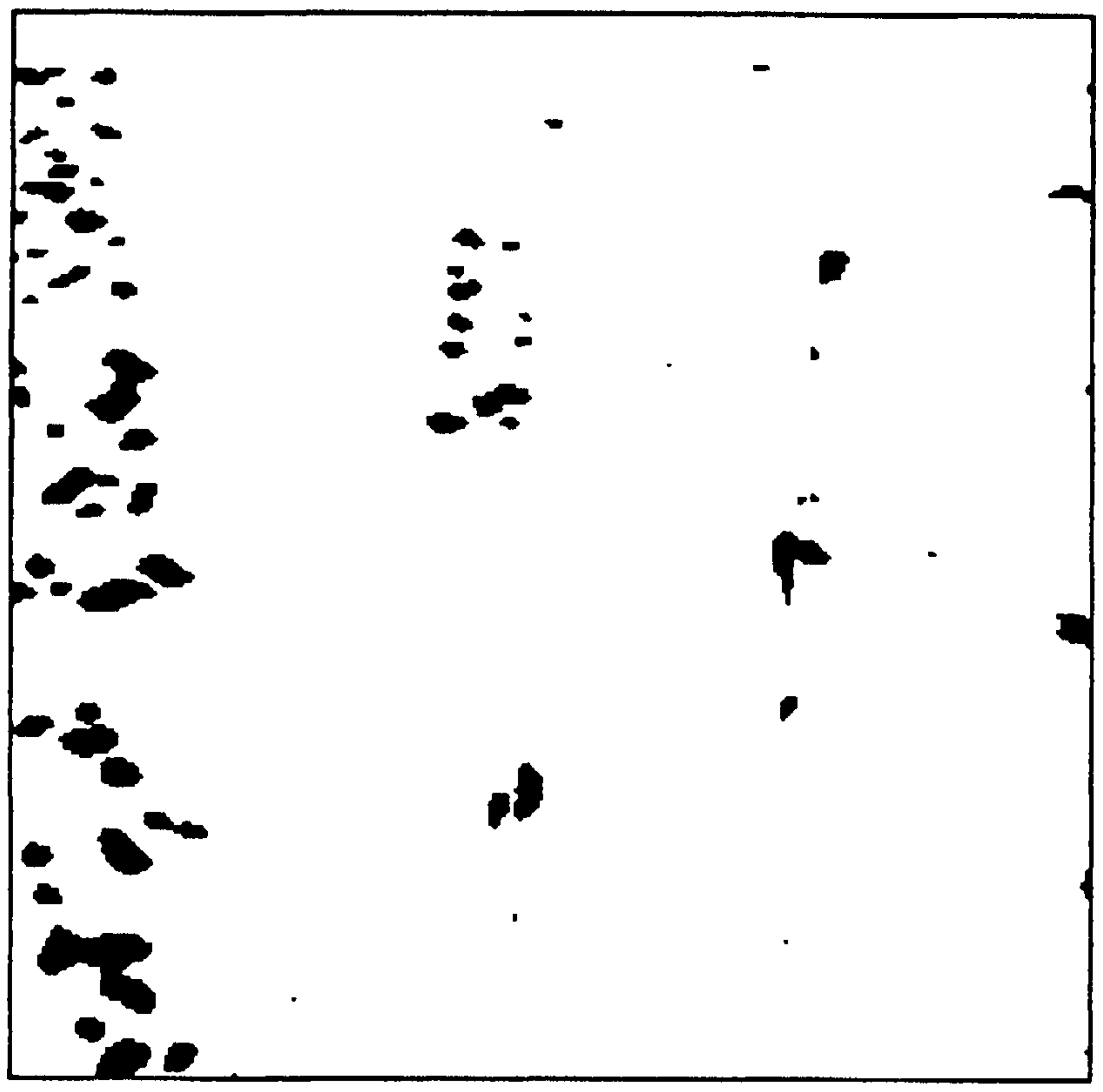


Fig. 6 The regions that are not matched by the inequalities and the "pulling effect".

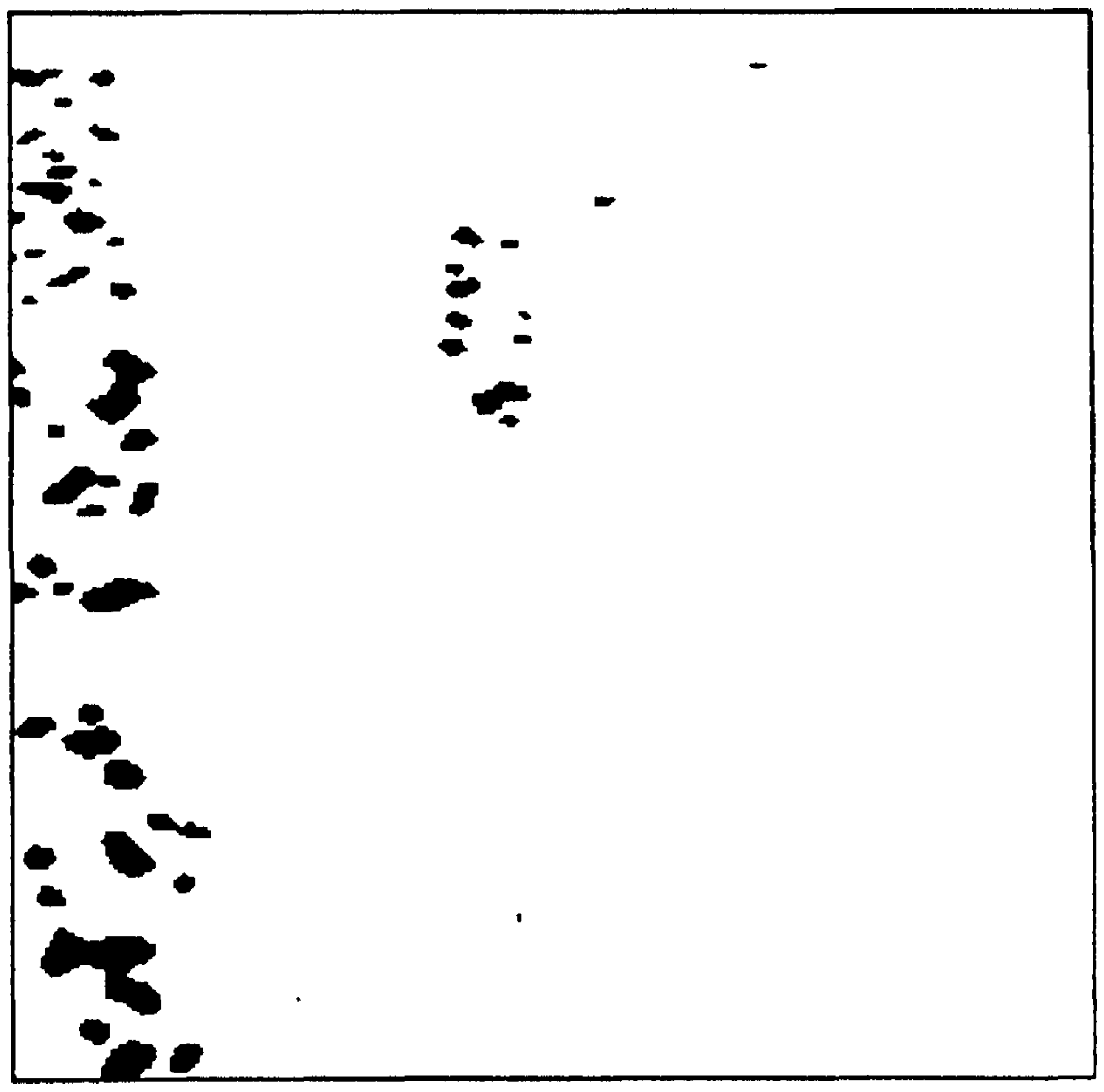


Fig. 7 The regions that are finally unmacthed.

ary, and the boundary points whose disparities are more than 2 pixels away from the peak value are discarded as noise. To the unmatched 75 regions, we further apply the method described in Section 5 to find the one-to-two correspondence relations and select matching points. Then the number of unmatched regions reduces to 57.

The histograms of the differences of area, perimeter and height of the regions between the two images are shown in Fig. 8. As evident in the figure, there is a strong correlation between the area difference distribution and the perimeter difference distribution. As a trial we removed perimeter from the matching algorithm, and the result showed that removing it did not influence the matching. This suggests that we use only a.rea and height as the descriptions for each region. Height is theoretically completely invariant between the two images, and actually the difference of height is no more than 2 pixels for nearly all regions, as can be seen from the figure.

The disparity map is shown in Fig. 9, with disparity encoded as brightness. The maximal and minimal disparities are 53 and 22 pixels, respectively.

The experimental results were generally very satisfactory. We consider that the processing time for matching can be further shortened. And a denser pattern image could improve the quality of the range data. More experiments are being carried out to make the algorithm applicable under various conditions.

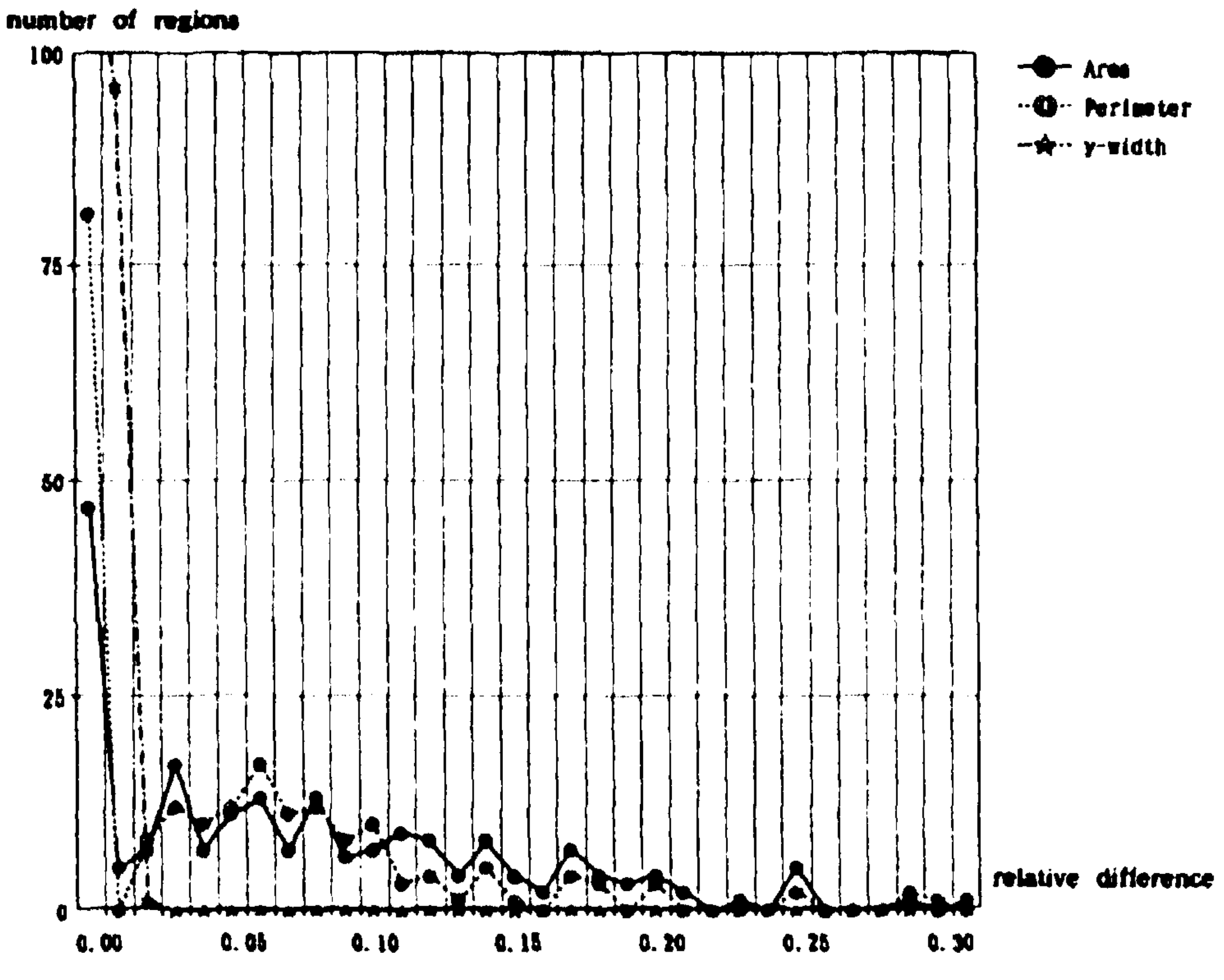


Fig. 8 A histogram of area, perimeter and height differences

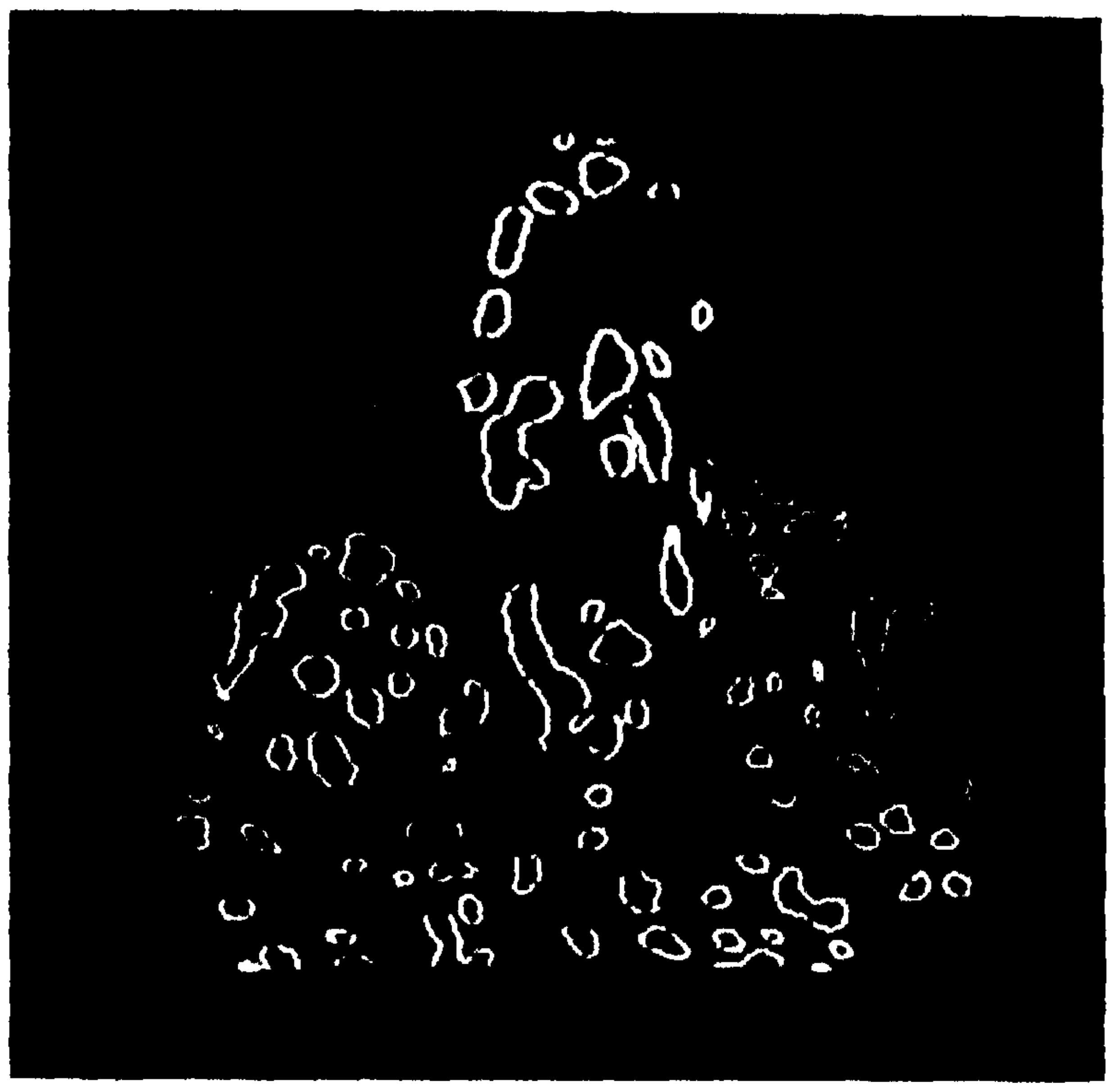


Fig. 9 The disparity map with disparity encoded as brightness

7. Conclusions

A region-based stereo algorithm has been proposed. The algorithm is original in that it matches large regions rather than segments or feature points. The region segmentation problem is overcome by projecting a random pattern onto the visual field. Regions are matched by their areas and heights that are relatively invariant between the two images. Disparity values are then measured along boundary points of matched regions. A special mechanism is developed to deal with regions that cross occluding boundaries. The algorithm yields fast, dense range data and is partcularly suitable for robotics tasks. It has been implemented to match pattern-projected real images. Experimental results show that the algorithm is efficient, reliable and tolerant of small vertical disparities. More experiments are being carried out to make the algorithm applicable under various conditions.

Appendix The Variations of Region Descriptions between Images

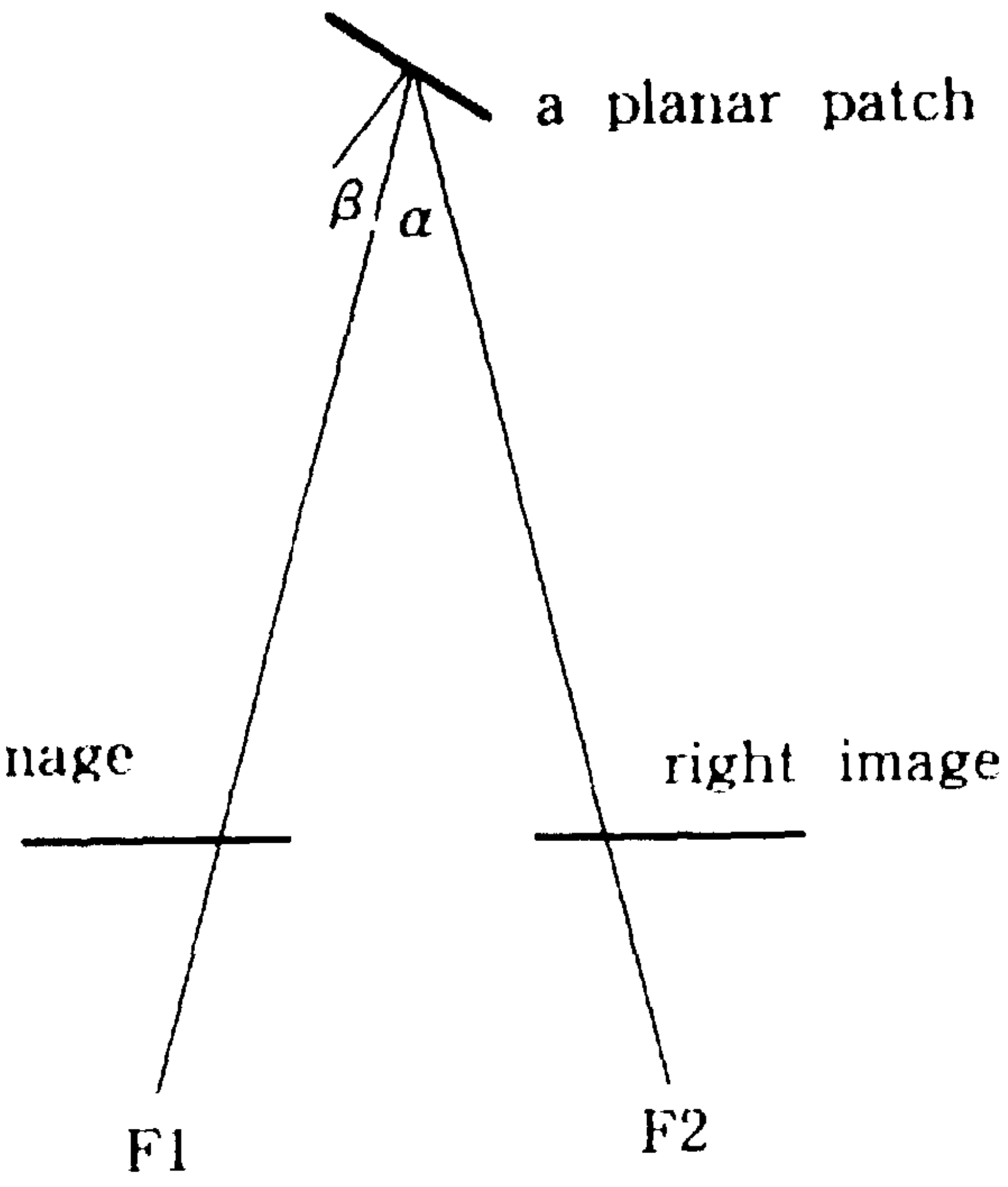


Fig. 10 Area variation of a vertical planar surface patch

Suppose that projection of the random pattern generates a bright spot on a surface in the scene. The surface patch can be approximated as planar if it is not very large. The variation in area between the left and right images is maximal if the planar surface patch is vertical with respect to the camera-centered coordinate system as shown in Fig. 10. The angle between the two lines of sight is α, and the angle between the normal of the planar patch and the left line of sight is β. If the area of the patch is a, then the area of its projection in the left image is approximately $a\cos\beta$, and that in the right image is approximately $a\cos(\alpha+\beta)$. The relative difference between them is defined as

$$\epsilon = \frac{a\cos\beta - a\cos(\alpha+\beta)}{a\cos\beta}$$

$$= 1 - (\cos\alpha - \sin\alpha\tan\beta).$$

Qualitatively speaking, ϵ increases as α does. α reaches maximum when the disparity reaches maximum. Suppose that α is 20° at the maximal disparity. Then

$$\epsilon = 1 - (\cos 20^\circ - \sin 20^\circ \tan\beta)$$

$$\approx 0.06 + 0.34\tan\beta.$$

When β is 45°, the relative difference is

$$\epsilon \approx 0.06 + 0.34\tan 45^\circ \doteq 0.4.$$

A region's boundary is composed of horizontal boundary points and vertical boundary points. Its perimeter is the number of the boundary points. When a surface patch is projected onto the two images, only the length of the horizontal boundary points varies, while that of the vertical boundary points does not. Thus the difference in perimeter between the two images depends on how many horizontal boundary points the whole boundary includes. For a space rectangle whose perimeter is $4l$, the relative difference is

$$e = \frac{(2l + 2l\cos\beta) - (2l + 2l\cos(\alpha+\beta))}{2l + 2l\cos\beta}$$
$$= \frac{\cos\beta - \cos(\alpha+\beta)}{1+\cos\beta}.$$

When $\alpha = 20°$ and $\beta = 45°$, $e \approx 0.16$.

The height of a region is invariant between the two images.

References

Ballard, D. and Brown, C. (1982) Computer Vision, Prentice Hall, Inc.

Barrow, H. and Popplestone, R. (1971) Relational descriptions in picture processing, Machine Intelligence 6

Echigo, T. and Yachida, M. (1985) A fast method for extraction of 3-D information using multiple stripes and two cameras, Proc. 9th Int. Joint Conf. on Artificial Intelligence, pp.1127 1130

Grimson, W. (1981a) From Images To Surfaces: A Computational Study of the Human Early Vision System, Cambridge, MA, MIT Press

Grimson, W. (1981b) A computer implememntation of a theory of human stereo vision, Phil. Trans. Roy. Soc, London, Vol. B292, pp. 217-253

Grimson, W. (1985) Computational experiments with a feature based stereo algorithm, IEEE Trans. PAMI, Vol.7, No.l, pp. 17-34

Marr, D. and Poggio, T. (1979) A computational theory of human stereo vision, Proc. Roy. Soc, London, Vol. B204, pp.301-338

Medioni, G. and Nevatia, R. (1985) Segment-based stereo matching, Computer Vision, Graphics and Image Processing, Vol.31, No.l, pp.2-18

Nishihara, H. (1983) PRISM: A practical realtime imaging stereo matcher, Proc. 3rd Conf. on Robot Vision and Sensory Controls, pp.121-129

Ohta, Y. and Kanade, T. (1985) Stereo by intra- and inter- scanline search using dynamic programing, IEEE Trans. PAMI, Vol.7, No.2, pp.139-154

Sato, K. and Inokuchi, S. (1987) Range imaging system utilizing nematic liquid crystal mask, Proc. 1st Int. Conf. Computer Vision, pp.657-661

Xu, G., Tsuji, S. and Asada, M. (1987) A motion stereo method based on coarse-to-fine control strategy, IEEE Trans. PAMI, Vol.9, No.2, pp.332-336