IMAGE TECHNIQUES FOR THE IDENTIFICATION OF DEPRESSIONS AND OTHER OBSTACLES IN AUTOMATED GUIDANCE OF ROVING ROBOTS

Malek Adjouadi
Department of Electrical Engineering
University of Hawaii at Manoa
Honolulu, Hawaii 96822

ABSTRACT

In this study, we analyze the problem of detection of depressions or drop-offs in the automated guidance of roving robots. The proposed approach is based on the principle that if one is too near a depression, one is bound to see new information which initially was occluded. To exploit this principle, two steps are undertaken. The first step involves the derivation of the correspondence process to allow the vision system to relate a location of interest in a sequence of frames. The second step involves the development of methods to detect and identify, in this location of interest, the occluded information.

RÉSUMÉ

Nous presentons dans cette étude une méthode visuelle pour la détection de dépression de terrain (creux, pente, bordure de pavé, etc) dans le but d'assurer sauf-conduit pour un robot autonome. Cette méthode est basée sur le principe que si on s'approche d'une dépression de terrain, on est apt d'apercevoir certains elements d'information qui auparavant n'etaient pas apparents. Pour exploiter ce principe, deux etapes d'action sont prises. Dans la première étape, nous etablissont le rapport qui existe entre une serie de photographes pour permettre au robot de reconnaitre une même location d'intérêt dans deux ou plusieurs de ces photographes. Ces photographes sont prises en succession et en s'approchant de cette location d'interêt. Dans la deuxième étape, nous décrivont les techniques necessaires pour identifier et extraire les éléments d'information qui caracterisent la presence d'une depression de terrain.

INTRODUCTION

Depressions or drop-offs constitute a serious problem in the automated guidance of roving robots. Unfortunately, the detection of depressions is also a complex image analysis problem. In the human vision system, many visual cues such as stereopsis, occlusion cues, context in the viewed scene, change in textural properties, etc., are all interpreted and integrated with relative ease to yield an almost effortless perception of what, in fact, is a complex perceptual task. In image processing, however, a computer implementation exploiting any one of the aforementioned cues becomes a complex information processing problem. Clearly, there is no simple way to solve this problem. In the approach proposed here, the aim is to extract the occluded information given a sequence of frames based on methods which allow for a relaxed image correspondence process between these frames. The two methods devised here make use of intensity profiles or pixel intensity distributions. The first method identifies primary cues which suggest the presence of a depression. The second method extracts the occluded information to confirm the presence of a depression.

Before we describe the approach for the detection of depressions, we first take a broad view of the automated guidance of roving robots. In this view, with the aim to extend beyond the ideal settings generally considered, we make an assessment of real-world scenes identifying pertinent problems towards enhanced guidance of roving robots.

SCENE INTERPRETATION PROCESS

Figure 1 illustrates a process of analysis and interpretation of real-world scenes. In this process, the first function of the vision system is to provide the robot with a safety path. To carry out this function, the vision system uses the first-pass evaluation process[1] which exploits the surface consistency constraint[2] by comparing the environment ahead of the robot with an initial environment which is already determined to be obstacle-free. Computer results of an implementation of this first-pass evaluation on outdoor scenes are shown in Figure 2. The second function of the vision system is to provide the robot with the needed additional information in the event where an object blocking the path of travel is detected, or if some landmark need to be identified. To carry out this second function, the objects the robot is likely to encounter are categorized, and their essential visual characteristics are identified. In the process of Figure 1, these categories are: (1) shadows (false alarm), (2) depressions, (3) upright objects, and (4) flat objects. The essential features characterizing the above categories are:

1. Shadow. A surface upon which a shadow is cast will preserve its intrinsic physical characteristics. This is due to the relatively uniform effect of shadow on the image gray level intensities.[3,4]

2. Depressions. In approaching a depression, one is bound to see new information that was previously occluded. We call this the occluded information.
3. Upright objects. Most man-made upright objects have straight vertical edges. The few other man-made objects together with the natural objects which do not have straight vertical edges can be distinguished by the manner in which they project onto the two-dimensional (2-D) image plane. An upright object projects in the 2-D image plane proportionally to the depth of field it occludes.[5]
4. Flat objects. Flat objects are affected by perspective. Also, a flat object projects in the 2-D image plane proportionally to its length.[5]

A methodology for guiding the robot through a given scene can be:

1. As an initial step, the vision system takes left, front and right images of the scene to acquire a wide-angle view. Each image is analyzed using the first-pass evaluation process. The results obtained from the three images are integrated to yield an optimal tracing of the safety path. We refer to this step as the initialization phase.
2. The robot takes the optimal path, and the vision system is directed to enter what we refer to as the motion phase. In this motion phase, the vision system processes images in the chosen direction of travel. The wide-angle view is no longer necessary, unless an obstruction is encountered and a new direction of travel must be taken. Moreover, the image taking process is a function of the range of the safety path. For example, if a path is obstacle-free for x steps, then an image may be taken every x_1 steps, where x_1 is the integer part of the fraction x/k and k=2,3,4,... depending on how large x is. In this phase, essential safety path cues such as path clear, obstacle ahead, turn left/right, are provided in real-time, and the timing of the vision system is such that it is always processing a few steps ahead of the robot.
3. If an object is found along the direction of travel, the vision system issues a warning signal to the robot and directs it to pause and takes another picture. The system then determines the range and extent of the object, and provides the robot with the necessary avoidance cues. If identification of the object is desired, the system enters the identification process. We refer to this step as the warning/identification phase. This phase can be carried either in a sequential mode or in a parallel mode. In the sequential mode, the processing task for which the primary cues can be found with the least amount of processing time is performed first. If the results are not conclusive, the next processing task is performed, and so on. In the parallel mode, all processing tasks are initiated simultaneously, and execution of these tasks ends as the first primary cues are determined.

In organizing all these information processing tasks to yield an integrated vision system, the following important points are considered:

1. Implementation of a methodology and a decision making process to insure that an information processing task is initiated only if primary cues justify its execution;
2. Allowing for concurrent processing in the development of these information processing tasks;
3. Allowing for one processing task to call upon another processing task if ambiguities arise in the image interpretation results.

We now describe the approach for the detection of depressions. This description starts with a presentation of the image correspondence process.

IMAGE CORRESPONDENCE PROCESS

For the image correspondence process, we derive equations for the correspondence, in both range and width, of any two image points, going from one frame to the next. But first, we need to define the mapping principles between the three-dimensional (3-D) real world and the 2-D image. Given Figure 3, using properties of similar triangles, measurements in width (W) and range (R) in the real-world environment are mapped in the (x,y) image coordinate system by the following relationships:

$$y_k=\frac{f[R(y_k)+h\tan\alpha]+f[R(y_k)\tan\alpha-h]\tan(\beta+\alpha)}{[R(y_k)+h\tan\alpha]\tan(\beta+\alpha)} \tag{1}$$

$$x_j-x_i=\frac{f+[R(y_i)-R(y_0)]}{fW(x_i,x_j)} \tag{2}$$

where h is the camera height; f is the camera focal length; α is the camera tilt angle; and $\beta = \arctan(L/h)$, with L being the range between the camera and the first point viewed by the camera. Note that if we let $\alpha = 0$, Eq. (1) takes the simple form

$$y_k=\frac{fh[R(y_k)-L]}{LR(y_k)} \tag{3}$$

1. Range correspondence: The objective here is to find how two points, $y^1_{\ell 1}$ and $y^1_{\ell 2}$, in the vertical axis of the first frame map into points, $y^2_{\ell 1}$ and $y^2_{\ell 2}$, in the second frame. The difference in the range of the two frames is r_1. The superscript, 1 or 2 denotes the frame identity. Coordinate $y^1_{\ell 1}$ is the point where the object area starts, and point $y^1_{\ell 2}$ is an arbitrary point a small distance away from $y^1_{\ell 1}$. The distance between these two points depends on the extent of the detected object. For simplicity, let us assume that the tilt angle α of the camera plane is zero. To find the mapping between points $(y^1_{\ell 1},y^1_{\ell 2})$ and $(y^2_{\ell 1},y^2_{\ell 2})$, we use Eq. (3) to obtain the following relationships:

$$y^1_{\ell 1}=\frac{fh[R(y^1_{\ell 1})-L]}{LR(y^1_{\ell 1})} \tag{4}$$

$$y^2_{\ell 1}=\frac{fh[R(y^1_{\ell 1})-r_1-L]}{L[R(y^1_{\ell 1})-r_1]} \tag{5}$$

where $R(y^1_{\ell 1})$ is estimated by the first-pass evaluation. Similarly, since $y^1_{\ell 2}$ is arbitrarily chosen, its range $R(y^1_{\ell 2})$ is easily determined, and thus we obtain

$$y^1_{\ell 2} = \frac{fh[R(y^1_{\ell 2}) - L]}{LR(y^1_{\ell 2})} \quad (6)$$

$$y^2_{\ell 2} = \frac{fh[R(y^1_{\ell 2}) - r_1 - L]}{L[R(y^1_{\ell 2}) - r_1]} \quad (7)$$

2. Width correspondence: In order to save processing time, it is useful to focus only on a certain width of the image where the object is detected. So, if we choose a segment of width delimited by x_{k1} and x_{k2} in the first frame, we need to find their corresponding projections in the second frame. This correspondence is found using Eq. (2). By substituting $R(y_{\ell 1})$ for $f + [R(y_i)-R(y_0)]$, for the first frame, we have

$$x^1_{k1}-x^1_{k2} = \frac{fW(x^1_{k1},x^1_{k2})}{R(y^1_{\ell 1})} \quad (8)$$

For the second frame, we have

$$x^2_{k1}-x^2_{k2} = \frac{fW(x^1_{k1},x^1_{k2})}{R(y^1_{\ell 1}) - r_1} \quad (9)$$

Using these relations, the vision system can now relate a location of interest in any two distinct frames separated by an arbitrary range r_1 (see Figure 4).

EXTRACTION OF THE OCCLUDED INFORMATION

Occluded information that is revealed in a subsequent frame can be perceptually very deceptive (see Figure 5). This results from the fact that if we cannot locate the same point of reference in the two frames then we may conclude that there is no relationship between these two frames. This stresses the importance of a reference point from which the system starts to look for occluded information; in our analysis this reference point is chosen in the proximity of the object as indicated by the first pass evaluation. Moreover, the detection of occluded information will disturb many of the physical relationships that previously existed between the various elements in the given scene. Utilizing these two ideas, we describe two simple methods to extract this occluded information.

Method 1

Step 1: Take a vertical scan from point $y^1_{\ell 1}$ to point $y^1_{\ell 2}$ to generate a vertical pixel intensity distribution or profile. We call this profile $P^1_{\ell 1 \ell 2}$. Similarly, generate a vertical pixel intensity profile, $P^2_{\ell 1 \ell 2}$, between $y^2_{\ell 1}$ and $y^2_{\ell 2}$.

Step 2: Locate the number of major disturbances (peaks) in the profile $P^2_{\ell 1 \ell 2}$ when compared with the peaks in $P^1_{\ell 1 \ell 2}$. If there exists a difference in the number of major disturbances in the two profiles, then this implies the detection of the occluded information. By major disturbance or peak, we mean a point in the profile whose gray level value exceed the value P_{max} given by the standard relation[6]

$$P_{max} = \mu_p + 0.5\,\sigma_p$$

for the same profile. Parameters μ_p and σ_p are the mean and standard deviation of the intensity profile, respectively.

This method is used by the integrated vision system for determining the primary cues.

Method 2

Step 1: Obtain a few horizontal scans between points x^1_{k1} and x^1_{k2} of the first frame starting at the vertical coordinate $y^1_{\ell 1}$. We call these intensity profiles $P^1_{k1k2}(i)$, where i denotes the i-th scan. Obtain similar scans from the second frame, using the coordinate $y^2_{\ell 1}$ as the starting point and x^2_{k1} and x^2_{k2} as the horizontal limits. We call these itensity profiles $P^2_{k1k2}(i)$.

Step 2: As in step 2 of the previous method, locate a difference in the number of major peaks appearing in $P^2_{k1k2}(i)$ when compared to $P^1_{k1k2}(i)$. If peaks are found, occluded information is detected.

This method confirms the results obtained using the previous method.

Computer examples of this procedure are illustrated in Figure 6. Note, that when no occluded information is found in this analysis, the object remains a potential obstacle.

CONCLUSION

We described in this study an approach for the detection of depressions. This approach is based on finding occluded information from a sequence of frames. We noted that if this occluded information was not found, the object in question remains a potential obstacle, and the appropriate processing task is initiated to identify its nature. An attractive feature of this approach is that the methods used for the extraction of occluded information allow for a relaxed image correspondence process. For example, if a given peak in the intensity profile on an initial frame is missing in the corresponding intensity profile of a subsequent frame, it becomes sufficient to look for a disturbance in these peaks going from one frame to the next. A computer implementation of this approach on real-world scenes produced very good results. Moreover, in the first part of this study, we discussed an image interpretation process which identifies pertinent problems towards enhanced guidance of roving robots.

REFERENCES

1. J.T. Tou and M. Adjouadi, "Computer Vision for Roving Robots," Proceedings of the Cannes Symposium, Cannes, France, 1985.
2. W.E.L. Grimson, "A Computational Theory of Visual Surface Interpolation," Philosophical Transactions of the Royal Society of London, Vol. B298, pp. 395-427, 1982.

3. H.G. Borrow and J.M. Tenenbaum, "Recovering Intrinsic Scene Characteristics from Images," in Computer Vision Systems, A.R. Hanson and E.M. Riseman, Eds., Academic Press, New York, 1978.
4. J.T. Tou and M. Adjouadi, "Shadow Analysis in Scene Interpretation," Proceedings of the 4th Scandinavian Conference on Image Analysis, Trondheim, Norway, June 1985.
5. M. Adjouadi, "Discrimination of Upright Objects from Flat-Lying Objects in Automatic Guidance of Roving Robots," SPIE's Symposium Southeast on Optics and Optoelectronic Systems, Orlando, Florida, March 1986.
6. D. Brzakovic, "Computer Based 3-D Scene Description from Texture," Ph.D. Dissertation, Department of Electrical Engineering, University of Florida, 1984.

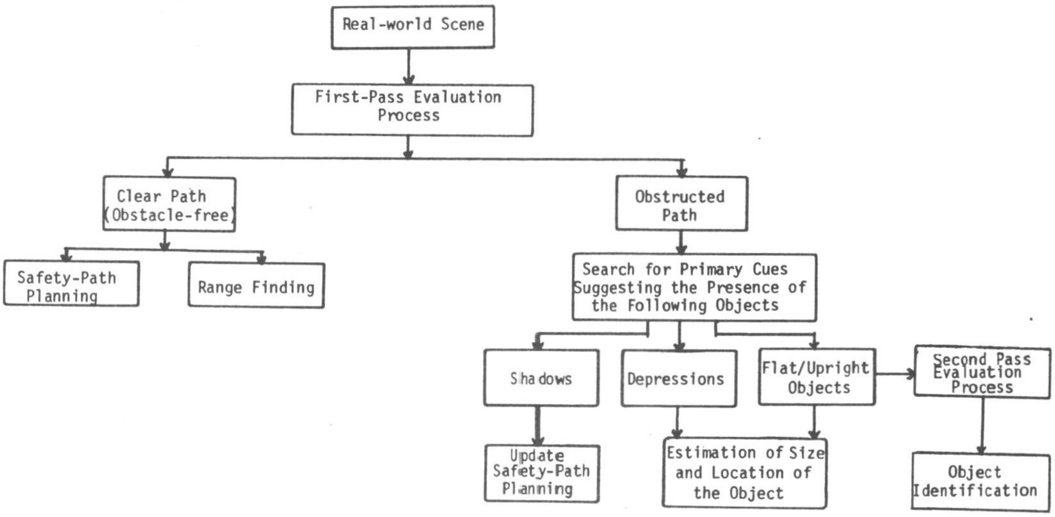


Figure 1. Process of Analysis of Real-World Scenes

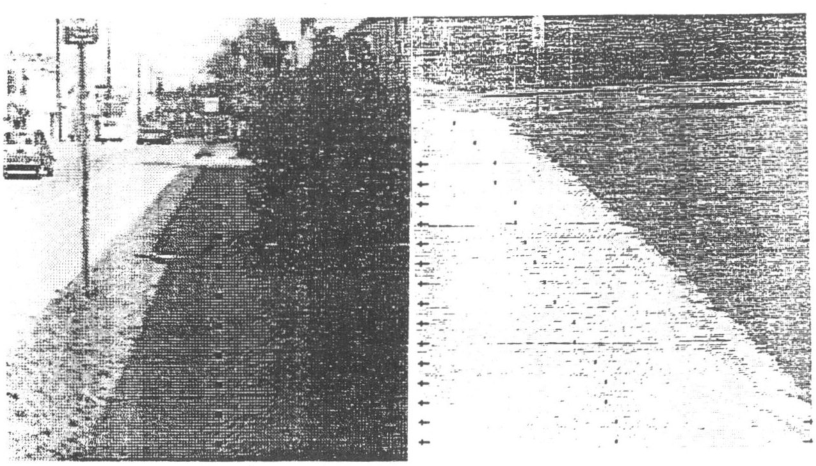
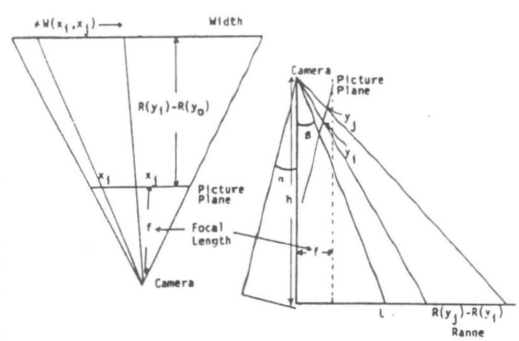


Figure 2. Results of the First-Pass Evaluation on Two Outdoor Scenes



(a) Mapping of Width (b) Mapping of Range

Figure 3. Mapping of Real-World Measurements onto the Image Plane

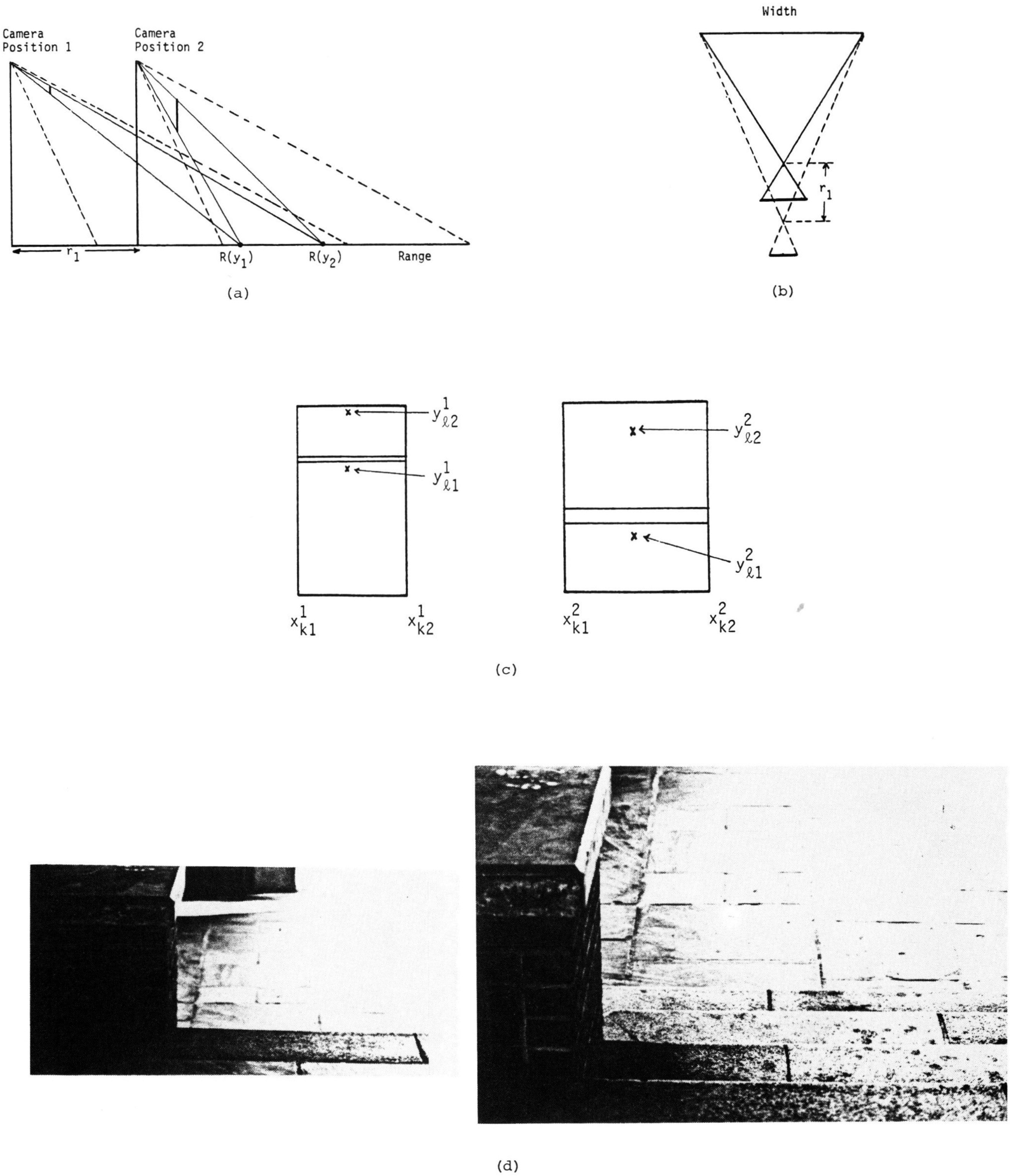
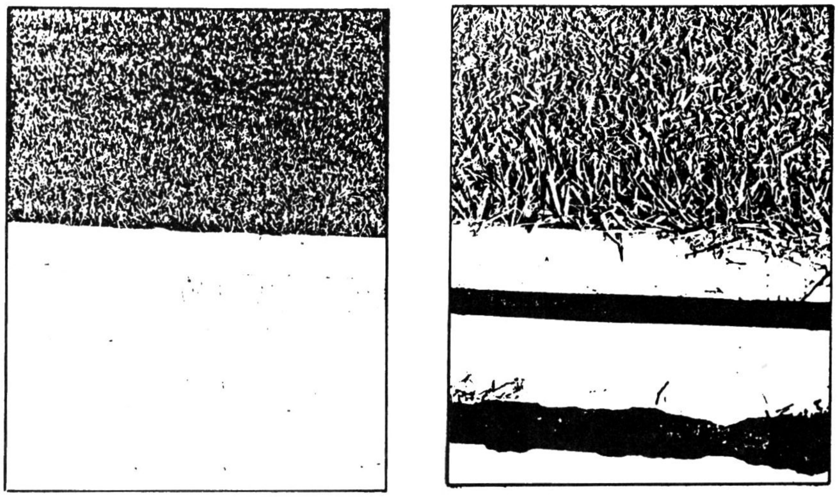
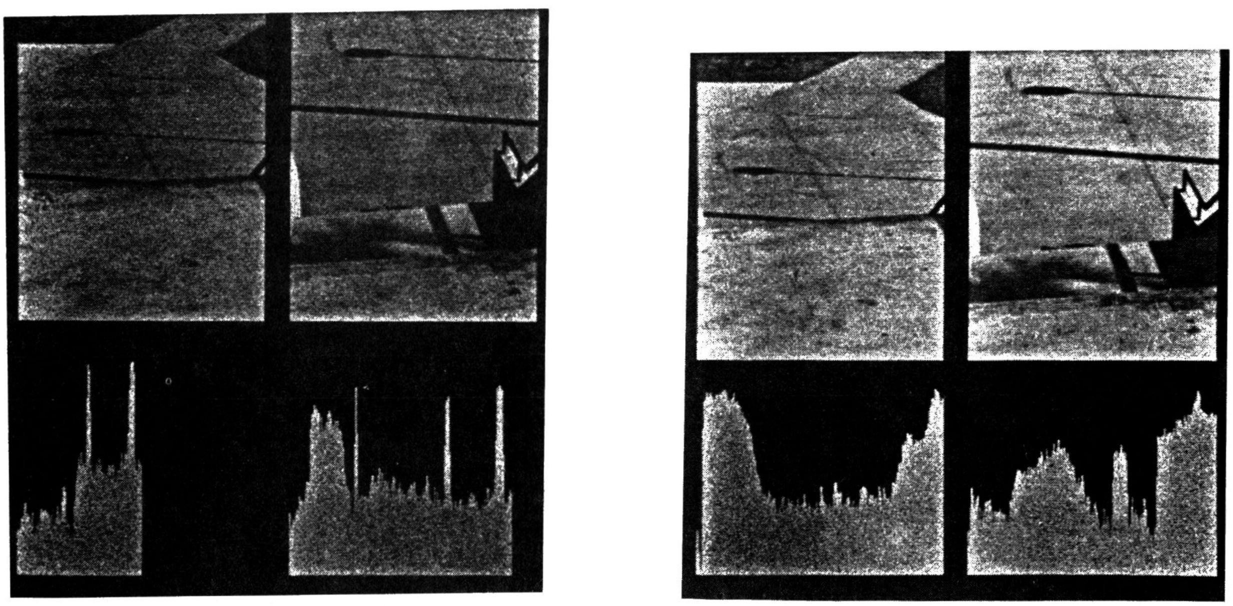


Figure 4. Image Correspondence for Extracting Occluded Information. (a) Correspondence in Range; (b) Correspondence in Width; (c) Mapping the Reference Points; (d) Mapping the Location of Interest.



(a) Input Image A (b) Closer Range of Input Image A

Figure 5. Need for a Point of Reference to Extract Occluded Information



(a) (b)

Figure 6. Extraction of Occluded Information. Input Images and their
(a) corresponding vertical intensity profiles with occluded information revealed in the close range frame;
(b) corresponding horizontal intensity profiles with occluded information revealed in the close range frame.