Department of
Veterans Affairs

Journal of Rehabilitation Research
and Development Vol. 29 No. 2, 1992
Pages 57–76

# A man-machine vision interface for sensing the environment

**Malek Adjouadi, PhD**
*Department of Electrical and Computer Engineering, Florida International University, Miami, FL 33199*

**Abstract**—This study describes a computer vision approach for sensing the environment with the intent of helping people with a visual impairment. The principal goal in applying computer vision is to exploit, in an optimal fashion, the information acquired by the camera(s) to yield useful descriptions of the viewed environment. The objective is to seek efficient and reliable guidance cues in order to improve the mobility needs of individuals with a visual impairment.

In this research direction, the following problems are identified and addressed: 1) the vision system design; 2) establishment of the mapping principles between the two-dimensional (2-D) camera images and the three-dimensional (3-D) real world; 3) development of appropriate imaging techniques for the interpretation of the 2-D images; and, 4) establishment of a communication link between the vision system and the user. The soundness of this research direction is assessed by means of a theoretical framework and experimental evaluations.

**Key words:** *artificial vision system, computer vision, guidance aid for blind, visual impairment.*

## INTRODUCTION

During the past half century, the needs of visually impaired individuals have been defined and explored across the spectrum, from the social to the scientific and technological. Notable advances have been made in the areas of social adjustment, vocational rehabilitation, and in the design of communication and learning devices. However, with the present scientific and technological breakthroughs, we are still reminded of the challenge put forth by Zahl, "A civilization with such skills should be able to develop guidance aids for the blind more knowing than the cane, more dependable than the dog" (p. 443) (1).

The design of guidance aids for the visually impaired can be pursued along two major directions: 1) applications of electromagnetic and sonic technologies to provide obstacle detection cues with, at best, limited information on the detected obstacles and the sensed environment in general; and, 2) application of computer vision toward optimal utilization of the acquired information of the cameras to yield reliable guidance cues and suitable descriptions of the viewed environment.

A majority of the research and development work on the aforementioned guidance aids in the past two decades pursued the first research direction. To pursue the second research direction requires the design of computer vision systems equipped with appropriate algorithms capable of intelligently and efficiently analyzing and interpreting real-world scenes. The work presented here follows the second direction.

### Review on guidance aids

As early as 1944, a National Research Council recognized that there was a desperate need for research and development of sensory devices to help visually impaired individuals in their mobility needs (1). However, even with today's space-age and information-age technology, providing practical means to aid the visually impaired in their search for relatively easy and efficient mobility is a problem that has yet to be resolved.

Early concepts that led to the design of guidance aids for the visually impaired can be classified into three main categories: 1) devices which make use of electromagnetic waves; 2) devices which make use of ultrasonic waves; and, 3) tactile vision devices, which transform simple visual information into tactile information (2–6).

Address all correspondence and requests for reprints to: Dr. M. Adjouadi, Department of Electrical and Computer Engineering, Florida International University, University Park Campus, Miami, FL 33199.

DOI: 10.1682/JRRD.1992.04.0057

The principle of electromagnetic and ultrasonic devices is based upon emitting electromagnetic or ultrasonic signals and detecting and decoding that portion of the signal that is reflected by encountered objects. Most of these devices are simple designs with varying degrees of practicality. Some of them are mere obstacle detectors; others, with some added sophistication, can provide range information and even some primitive features of the object, such as texture. Certainly, these devices have their practical uses, but to serve as guidance aids they must first overcome the following limitations: 1) environmental sensing capability is limited to short-range objects, small or thin objects, or obstacles such as curbs which may not be detected; 2) it is difficult to make any spatial judgment about the detected object(s) from the reflected signal; 3) not all objects reflect the transmitted signal, and weak reflections may not be detected due to outdoor noise; and, 4) directionality is often reduced to going from one detected obstacle to another.

The tactile vision devices convert visual information into tactile information. The visual information is mapped into an array of vibro-tactile stimulators. This is reduced, however, to an ON and OFF type of information wherein the ON stimulators indicate the presence of the object and the OFF stimulators characterize the background. These devices are still in the experimental stage and suffer from the following problems: 1) biophysical difficulties related to the limited form-sensing and spatial resolution of the skin, and to the crosstalk between stimuli (7); 2) complex images yield complex tactile patterns which exceed the perceptual integration capability of the skin, thus making the tactile patterns extremely difficult to interpret; 3) there is no distinct way to extract the range information; and, 4) only closer objects may be recognized.

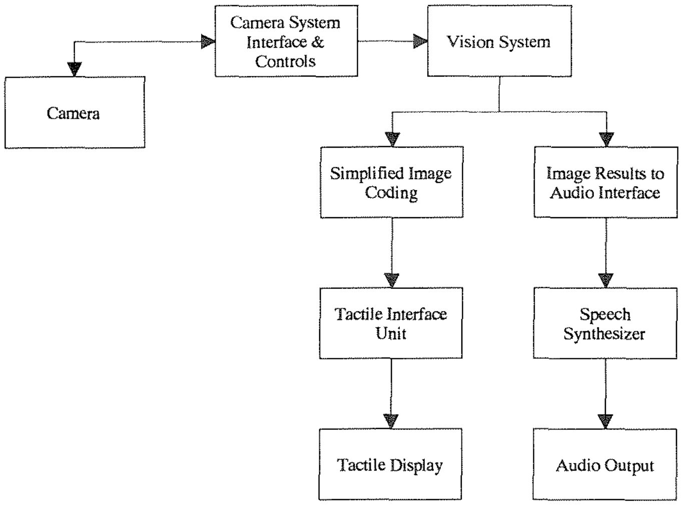


**Figure 1.**
Computer vision system.

The considerable progress made in computer vision and image processing applications, together with the improvements in size and speed of computers, have led to a new research direction: the computer vision approach. Along this line of research, interesting studies have been reported (8–13). Although only Deering's work (9) deals specifically with guidance for individuals with a visual impairment, the central theme of all these studies is still that of processing 2-D images of real-world scenes. Under the constraints and ideal settings assumed, there are many contributions we can attribute to these studies and, with continued research efforts, these studies could prove useful to computer vision-based guidance systems.

## METHODS

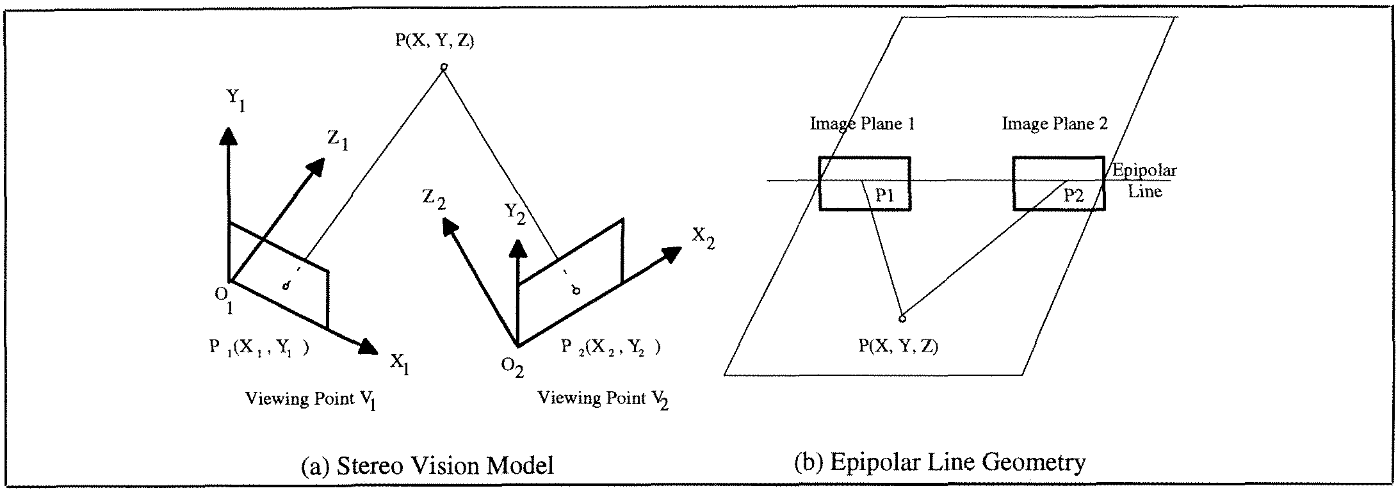
### Computer vision approach

In this approach, the main effort is in fulfilling the requirements of 1) directional guidance—planning or tracing a safety path; 2) *orientation information*— establishing the spatial relationships between the user and objects in the scene; 3) depth information extraction—constructing a depth map (a 3-D interpretation of the viewed scene); 4) obstacle detection and avoidance—warning of obstacles and providing avoidance cues; and, 5) object identification—identifying those objects deemed important in the guidance process.

*The vision system design*

A computer vision system design is illustrated in **Figure 1**. The main components of this system are:

1. Two Charge Coupled Device (CCD) cameras to serve as the sensing devices of the real world (the use of two cameras is required in our efforts to implement special motion vision and stereo vision algorithms).
2. A microcomputer system equipped with vision algorithms to analyze and interpret the 2-D images. These vision algorithms, which are organized in a modular structure that lends itself to parallel processing, tackle the following problems: (*a*) planning of a safety path; (*b*) detection of depressions or drop-offs; (*c*) discrimination of upright objects from flat objects; (*d*) identification of shadows (false alarm); and, (*e*) identification of relevant objects (e.g., staircase, crosswalk, curb, doorway, etc.).
3. An output interface unit with audio and tactile features to relay to the user a suitable description of the



(a) Stereo Vision Model

(b) Epipolar Line Geometry

**Figure 2.**
Image projections in stereo vision.

real world. In the audio unit, the concept of speech generation from a set of digitized words is considered. The auditory information, in this case, is separated into fixed format sentences such as "Path is clear," "You may turn left/right," and into variables (X,Y) which will provide the range information. The results will be, "Path is clear for X steps," "You may turn left/right after Y steps." In the tactile unit, only the safety path is displayed in order to eliminate the tactile vision problems discussed earlier, and the range information can be conveyed in the form of a coded signal which represents the number of walking steps. This is particularly important in the case of individuals who are both visually impaired and deaf.

*Mapping principles between the 2-D image and the 3-D real world*

In this section, a stereo vision method based on the theory of scaled-space filtering is investigated. The goal is to recover the depth information from the disparity between a stereo pair of images. An observation is made on the motion vision method. It is worthy to note that a new approach to the problem would be to deal with so-called 2½-D augmented image given the commitment that a depth map of the viewed scene can be acquired in real-time (14). Undoubtedly, scene interpretation would be enhanced if the more revelatory augmented images were used.

*Depth perception using stereo vision*

This method applies the theory of scaled-space filtering: stereo matching of features of interest proceeds from coarse image scale (large $\sigma$) to fine image scale (small $\sigma$) to determine the disparity measure accurately. We recall that the basic principle in stereo vision is to measure, through a correspondence process, the disparity that exists between a stereo pair of images; this disparity is a function of the depth information we are seeking. The image scale concept is a function of the space parameter, $\sigma$, of the Gaussian function:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

A physical point, P(X,Y,Z), in a scene projects onto $p_1(x_1,y_1)$ on the first image plane, and onto $p_2(x_2,y_2)$ on the second image plane, as shown in **Figure 2**. The disparity, $d(x_1,y_1)$, is the distance between the two corresponding image points, $p_1(x_1,y_1)$ and $p_2(x_2,y_2)$, when two image planes are superimposed. By simple geometry, it can be shown that depth, $Z(x_1,y_1)$, is inversely proportional to the disparity, $d(x_1,y_1)$, provided a viewer-centered coordinate system is used.

$$Z(x_1,y_1) = \frac{Bf}{d(x_1,y_1)}$$

where $B$ is the distance between the two cameras, and $f$ is the focal length of the cameras.

Once the disparity, $d(x_1,y_1)$, is found, the depth, $Z(x_1,y_1)$, can be computed exactly, since the product, Bf, is a known constant. Computing depth is therefore not an issue: the real problem is in determining the disparity.

The two-parallel-camera model, as illustrated in **Figure 2**, was designed to make the epipolar line parallel to the horizontal scan lines of the two image planes. A

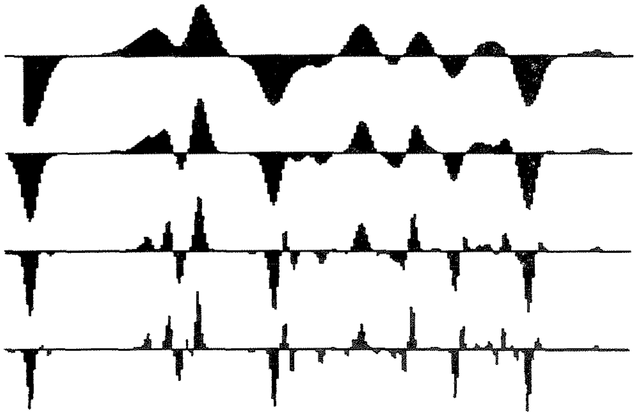


**Figure 3.**
Example of 1-D scale-spaced filtered intensity profile.

viewer-centered coordinate system and positive image planes are assumed. Two cameras were mounted rigidly, with their optical axes parallel to each other, pointing at positive Z direction, separated by a distance, B, in the X direction. With this camera configuration, the vertical disparity is zero, and the horizontal disparity is non-zero and inversely proportional to depth.

To determine the disparity using the theory of scaled-spaced filtering, the following steps are considered:

1. The stereo pair of images are first filtered using Laplacian of Gaussian operators (in this approach, only the 1-D Laplacian of Gaussian are used on each image record ($y_i$)

$$\nabla G(x) = \frac{1}{2\pi\sigma^4}(-x)e^{-\frac{x^2}{2\sigma^2}}$$

   where Δ denotes the first order Laplacian operator. The filtering (smoothing of the image) effect reduces the likelihood of feature mismatching.
2. A segmentation process is applied to determine for each image record a sequence of image features as related to peaks (maxima) and valleys (minima) of the 1-D filtered intensity profiles obtained in step 1. The peak location of an image feature signifies the feature's location, and the left and right valley locations signify the feature's left and right boundaries (**Figure 3**).
3. The features found in step 2 are then matched from coarse scale to fine scale to produce a set of scaled disparity maps which are combined into a composite multiscaled disparity map.

In the application of this method, it is worthy to note that (*a*) matching at various scales can interactively reinforce the confidence of matching; (*b*) matching at the coarser scales provides a more constrained search space for the matching at the finer scales (consequently, many false targets can be excluded); and, (*c*) the output of multiscaled matching is a full description of the analyzed scene, and this may help the interpretation of the sparse depth data to reconstruct the 3-D scene.

An example of the results obtained using this method is shown in **Figure 4**. The accuracy of these results varied from 80 percent to 99 percent depending on the complexity of the scene. For visual appreciation, the depth information is displayed as a function of brightness. Closer

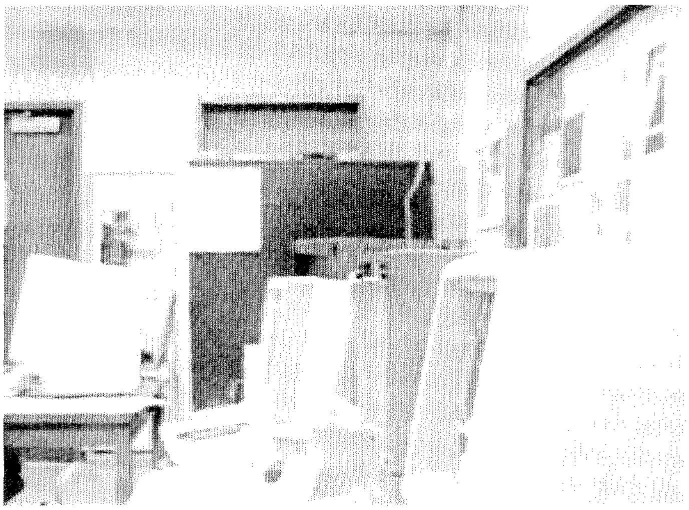


**Figure 4a.**
Depth information extraction using stereo vision: Input image.

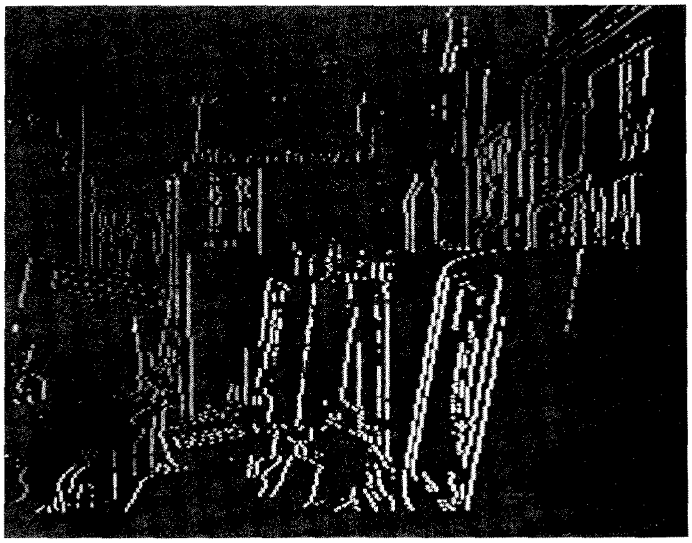


**Figure 4b.**
Depth information extraction using stereo vision: Results.

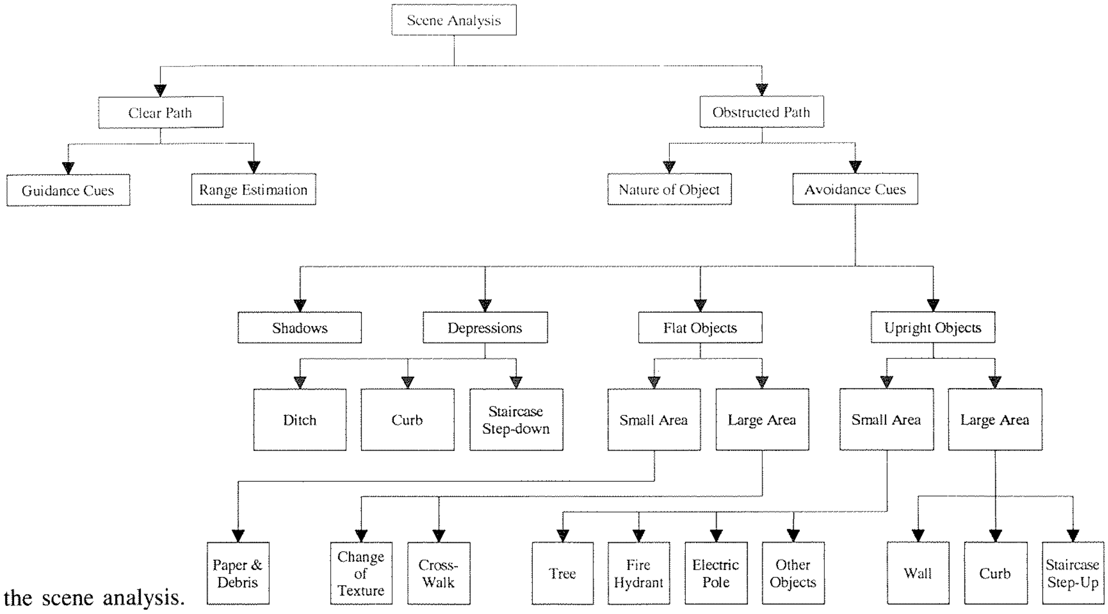


**Figure 5.**
Structure of the scene analysis.

objects appear brighter in **Figure 4b**. Adjouadi and Zhaing present a detailed description of this approach (15).

Observation

It is important to point out that another avenue to the recovery of the depth information is the exploitation of the motion vision principle. Simply stated, the basis of motion vision is the functional relationship that exists between the motion of the observer (the user) and the induced spatial and temporal information changes in a sequence of images. These information changes are functions of both the motion of the observer and the depth map of the scene, both of which can be determined under certain fairly broad conditions (16,17).

## Image analysis and interpretation

Before the various imaging techniques are detailed, a structure of the real-world domain that the vision system is to analyze is presented, and the organizing principles to tackle such a domain are emphasized. Following this, an integrated system incorporating all the imaging techniques is described.

### *Domain of analysis and organizing principles*

Before one can develop software modules for the interpretation of scenes, one must first ascertain the essential characteristics of scenes that the vision system is to exploit. A structure may then be established based on a logical order in which the various modules ought to be performed. In this structure, as illustrated in **Figure 5**, the vision system first determines whether the path of travel is obstacle-free or obstructed. If it is obstacle-free, the system will trace a safety path and estimate its range. If the path is obstructed, the system will either provide avoidance cues or, at the request of the user, determine the nature of the object. The methodology is as follows:

1. As an initial step, the vision system takes left, front, and right images of the viewed scene to acquire a wide-angle view. Each image is analyzed using the first-pass evaluation technique described under *Scene analysis for safety path planning*. The results derived from the three images are integrated to yield an optimal tracing of the safety path. This step has been named the "initialization phase."
2. The user decides on a given direction of travel provided by the safety-path tracing, and the vision system is directed to enter the "walking phase," wherein the vision system processes images in the chosen direction of travel. The wide-angle view is no longer necessary, unless a major obstruction is encountered and a new direction of travel is to be chosen. The image-taking process, in the final implementation, should be a function of the depth (length) of the safety path and should take into consideration possible deviations from the path of travel. In this phase, essential safety path cues such as *safe step, obstacle ahead, turn left/right,* can be provided in real-time. In fact, an implementation of this procedure on a roving robot has been done successfully. It took from 20 to 38 seconds (depending on the complexity of the viewed scenes) between the execution time to the actual movement of the roving robot toward an assigned destination following the safety path generated using the

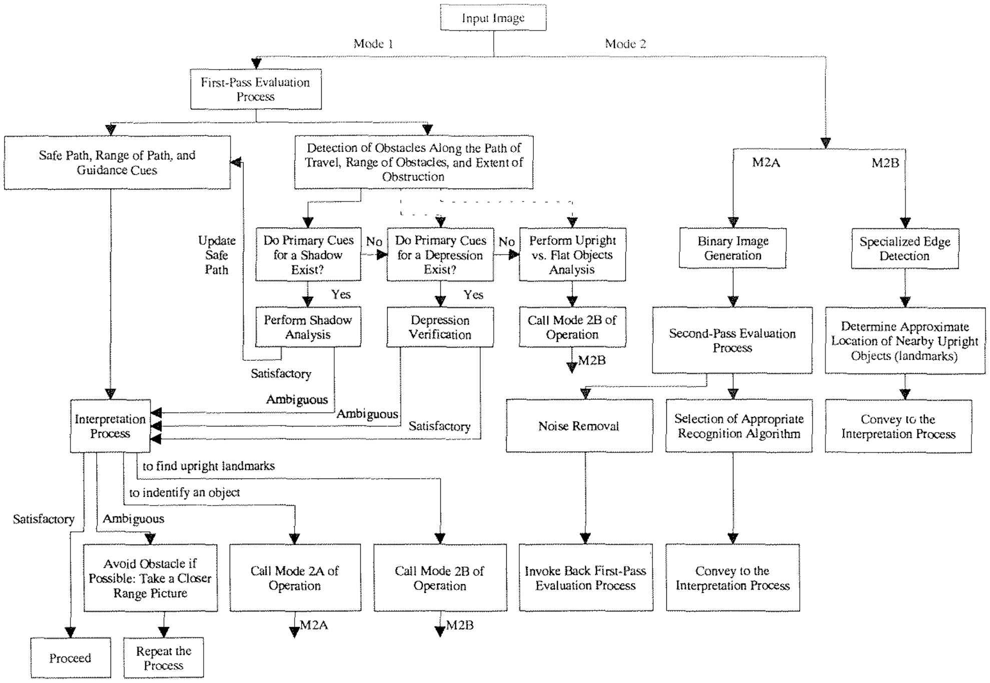

Input Image
Mode 1
Mode 2
First-Pass Evaluation Process
Safe Path, Range of Path, and Guidance Cues
Detection of Obstacles Along the Path of Travel, Range of Obstacles, and Extent of Obstruction
M2A
M2B
Update Safe Path
Do Primary Cues for a Shadow Exist?
No
Do Primary Cues for a Depression Exist?
No
Perform Upright vs. Flat Objects Analysis
Binary Image Generation
Specialized Edge Detection
Yes
Yes
Perform Shadow Analysis
Depression Verification
Call Mode 2B of Operation
M2B
Second-Pass Evaluation Process
Determine Approximate Location of Nearby Upright Objects (landmarks)
Satisfactory
Ambiguous
Ambiguous
Satisfactory
Interpretation Process
Noise Removal
Selection of Appropriate Recognition Algorithm
Convey to the Interpretation Process
to find upright landmarks
to indentify an object
Satisfactory
Ambiguous
Avoid Obstacle if Possible: Take a Closer Range Picture
Call Mode 2A of Operation
Call Mode 2B of Operation
Invoke Back First-Pass Evaluation Process
Convey to the Interpretation Process
M2A
M2B
Proceed
Repeat the Process

**Figure 6.**
Structure of the integrated vision.

above-described first and second steps. Close to 70 percent of the processing time is spent reading the image file and writing the results back onto the system's disk for viewing purposes. The input image, which is divided into sub-images of 10 records each (the number of records in an image is 480), could be processed in parallel—a processor for each sub-image would constitute the ideal case. The issue here is the cost/performance ratio. Also, simple image data compression schemes can be used initially to reduce the image size by a factor yielding minimal effect on the image data information, resulting in more processing time saved.

3. If an object is found along the direction of travel, the system issues a warning signal to the user and asks him/her to pause. The system then provides the necessary avoidance cues. If identification of the object is desired, the system enters the identification process. This step has been named the "warning/identification phase." The object is extracted from the background and a preliminary description is provided in an additional 10 to 15 seconds. The concept of parallelism noted in step 2 applies here as well.

*The integrated vision system*

The functional structure which links the various imaging techniques to yield the integrated vision system is illustrated in **Figure 6**. The first function of the integrated vision system is to trace a safety path. To carry out this function, the system uses the first-pass evaluation technique which is devised to exploit the surface consistency constraint. This constraint implies that a physical surface with a given orientation is continuous due to the coherence of matter and, consequently, will exhibit uniform reflectance. In a 2-D image, this fact translates proportionally into a surface with consistent gray-level intensities given the well-established linear relationship that exists between brightness in the image (irradiance) and brightness in the real world (radiance). In this system, the constraint is used by the first-pass evaluation technique to plan a safety path by comparing the environment that is ahead with an initial environment that has been determined to be obstacle-free.

The second function of the integrated vision system is to provide needed additional information in order to enhance the interpretation of the viewed scene. This additional information is accessible either directly, via the user's command, or automatically using the first-pass evaluation. The direct access mode is necessary when the user desires a primitive description of a detected object, or in extracting upright landmarks to help locate such things as a bus stop, the corner of a building, or a doorway. The automatic access mode is carried out if an object blocking the path of travel is detected by the first-pass evaluation. A key issue of the automatic access mode is having the information processing tasks organized in an efficient way. To

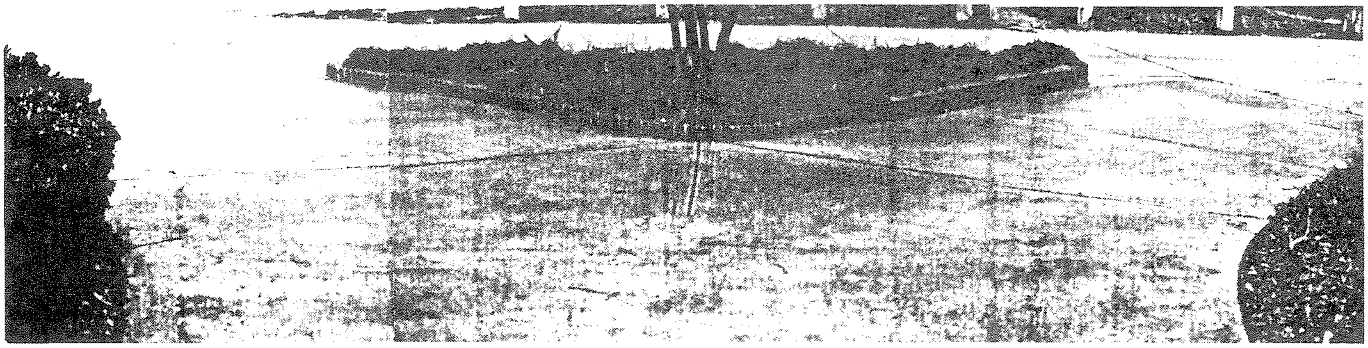


**Figure 7a.** Wide-angle view input image.

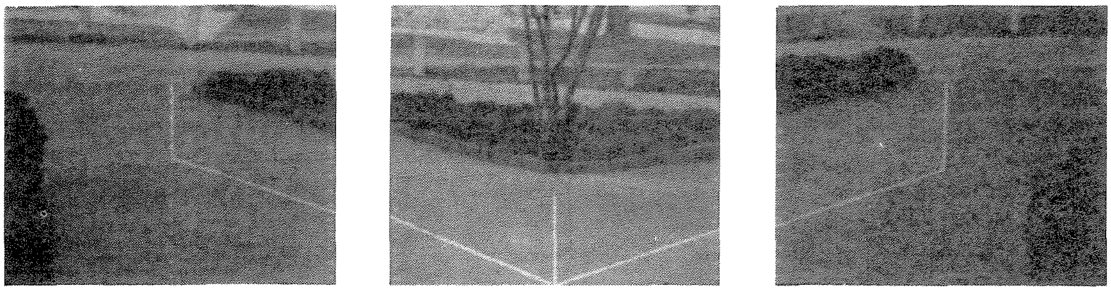


**Figure 7b.** Wide-angle view results.

do so, a decision-making process is devised based on the principle that the vision system is to verify the identity of an object only if some primary cues suggest its existence. When the decision-making process fails to produce conclusive results, the object is declared an obstacle (for safety purposes) regardless of its real nature. To complete identification in this case, a close range image of the object is necessary.

*Scene analysis for safety path planning*

Three image techniques used in conjunction with the safety path planning are described below.

Wide-angle view technique

Based on the importance of human peripheral vision, the wide-angle view technique is devised so that more information of the surrounding environment is gathered. The result, as shown in **Figure 7**, is an enhanced path tracing and guidance process.

Two practical steps constitute the wide-angle view technique. One is the acquisition, by the vision system, of a wide-angle view by taking left, front, and right images of the viewed environment. The second, and more important, step is the integration of the safety path results obtained from the images which are processed independently by the first-pass evaluation. This integration process yields safety path results which allow the user, when necessary, to change the direction of travel in an optimal way. The configuration and the integration process of this technique are described in **Appendix A**. An implementation example of this technique is shown in **Figure 7**.

First-pass evaluation technique

A basic assumption made in the first-pass evaluation is that the immediate area (the length of one step in the direction of travel) of the initial position of the user is safe or obstacle-free. This assumption, together with the surface consistency constraint, constitutes the core of the first-pass evaluation technique. The first step in the implementation of this technique consists of partitioning the image to be processed by a virtual grid whose unit is a cell containing 10×10 pixels or picture elements, as seen in **Figure 8**. This partitioning scheme is used to facilitate image processing and for future implementation applying parallel processing.

Based on a concise scheme which discriminates the cells as a function of their gray-level information, the first-pass evaluation technique performs the following analysis:

1. **Walking straight ahead.** This step is performed to plan a safety path in the direction of travel.

2. **Initial information about the object.** This step is performed to provide initial knowledge on the extent of the object, and to determine whether or not it can be avoided.

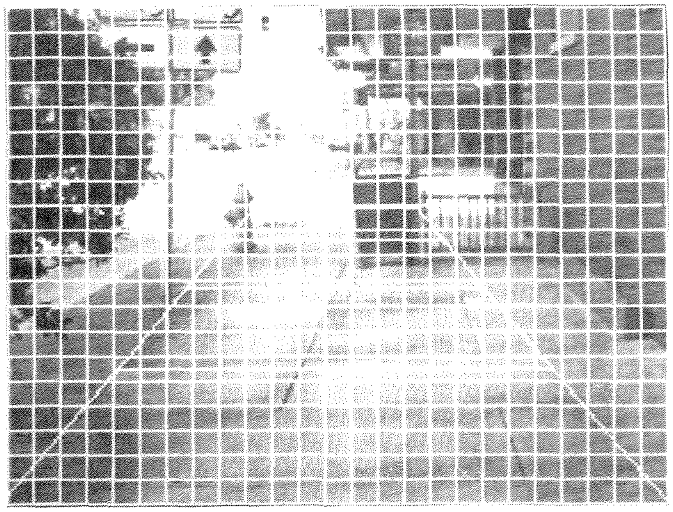


**Figure 8.** Partitioning scheme of the first-pass evaluation.

3. **Making a left or a right turn.** This step is performed when an obstacle is detected in the direction of travel. The two types of turns considered are (*a*) a turn to avoid a large obstacle, and (*b*) a turn to avoid a dead end. The distinction between condition (*a*) and condition (*b*) in this case is made through the nature of the path tracing generated by the first pass-evaluation process. The audio message is "Turn left/right after X steps." X is the range of the path provided by the algorithm described under *Depth perception using stereo vision*.

4. **Permissible left or right turns.** This step is performed to determine if a left or a right turn is permissible for the purpose of changing the direction of travel. The resulting audio message is, "You may turn left/right after X steps and the clearance is Y steps wide."

Experimental results of the first-pass evaluation using outdoor scenes are shown in **Figure 9**. In these results, we should indicate that each safety marker, generated by the vision system for visual evaluation, is a nonlinear function of depth.

Extensions of this first-pass evaluation technique are possible. For example, if a more elaborate scene analysis is desired, one could implement the perspective effect on the cells themselves. Also, to suit environments with curved paths as illustrated in **Figure 9d**, a variation of the straight-ahead analysis can provide the desired path. In this case, however, a least-squares approximation is required to estimate the direction of the path piecewise.

### Second-pass evaluation technique

The primary objective of the second-pass evaluation technique is to provide a primitive description of the object(s) detected by the first-pass evaluation. This primitive description is obtained from a segmented image. Image segmentation is generally used in the initial stages of image processing to highlight and extract object features from the background in order to facilitate image interpretation or object identification.

The central theme in image segmentation is the determination of the proper threshold which will separate the sought after object(s) from the background. Segmentation, in this case, is a process based on the gray-level variation that exists between the initial area assumed obstacle-free and the object area.

The gray-level threshold used to separate object from background is computed as follows:

$$T_s = \frac{1}{2[G_r + G_{obj}]}$$

where $G_r$ is the optimal regional average gray-level defined over an obstacle-free area in the image, and $G_{obj}$ is the gray-level average of an area within the object area. The images considered here have 256 (0–255) gray-levels. Recall that the presence of the object has been determined by the first-pass evaluation. With threshold $T_s$ obtained, we perform one of the following simple steps to extract an object from the background.

1. If $G_{obj} > G_r$, all points in the image whose gray levels exceed $T_s$ are set to 255, all others are set to zero.
2. If $G_{obj} < G_r$, all points in the image whose gray levels are less than $T_s$ are set to 255, all others are set to zero.

These two conditions insure extraction of the object from the background in the same fashion regardless of whether the object is lighter or darker than the obstacle-free area.

To save time, the focus is placed only on the area starting from a point near where the first-pass evaluation has indicated the presence of the object. From this point on, an n×n virtual grid is superimposed over the remaining area of the image. The unit cell of the grid which is still a 10×10 array is denoted by $C_k$, where $k=1, 2, \ldots, n^2$ denotes the number of the cell. Parameter $G(C_k)$ denotes the average gray-level of the cell, which in this case ranges from 0 for a background cell to 255 for an object cell. The objective here is to assess the manner by which the object is overlaid on the grid given the values $G(C_k)$, $k=1, 2, \ldots, n^2$. A primitive description is then obtained from this overlay. The procedure of the second-pass evaluation is described in **Appendix B**.

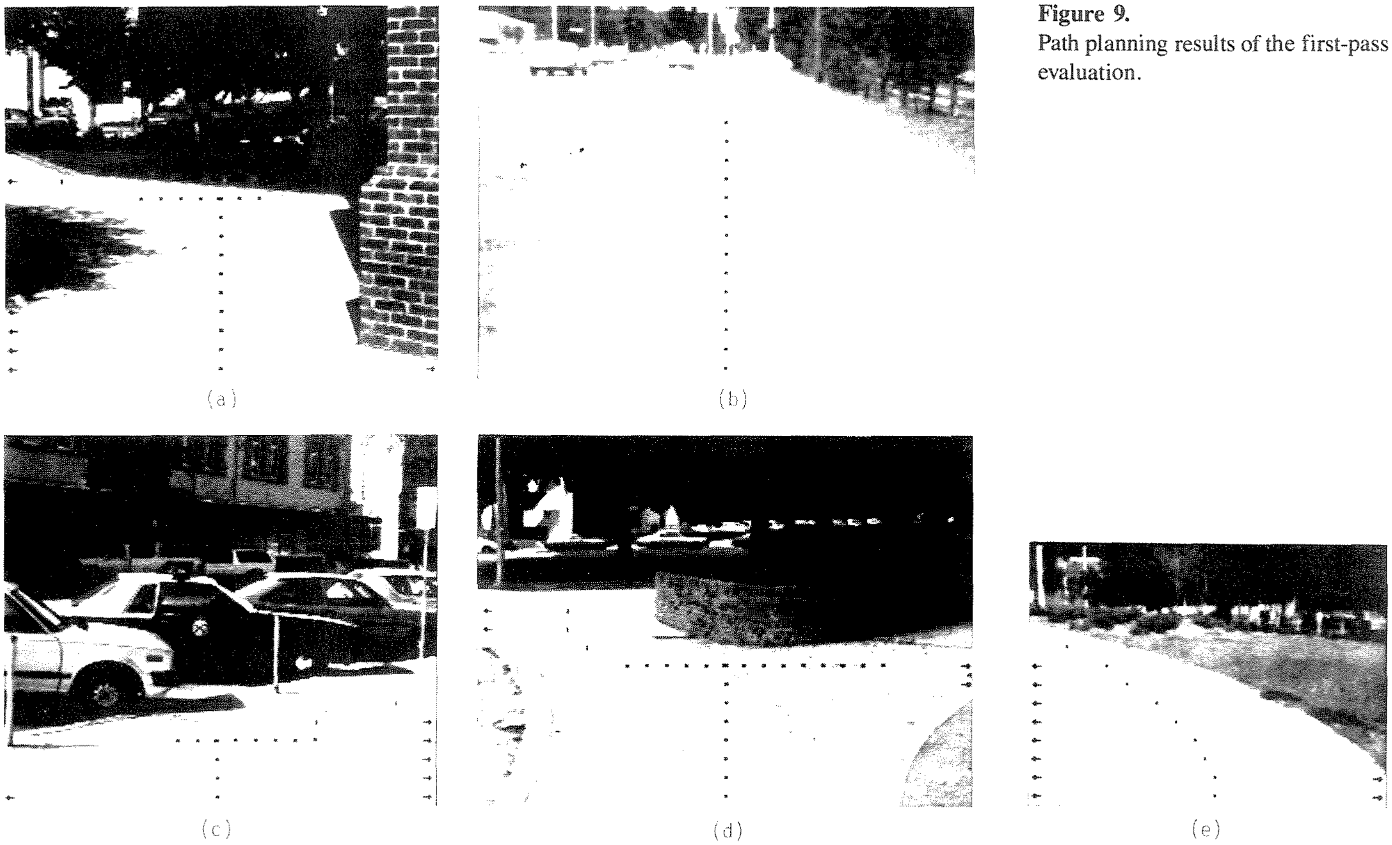


**Figure 9.**
Path planning results of the first-pass evaluation.

A technique such as this can be extended to include other possible interpretations depending upon the objects considered in the application. For example, if for the horizontal overlay ($H_C$), the $H_Cs \neq 0$ alternate with the $H_Cs=0$, this could serve as an indication of the presence of a staircase or a striped crosswalk. This last point is made to indicate that the primitive description provided by the second-pass evaluation can be used as a feature for initiating the proper recognition algorithm which will then identify the object. Results of a computer implementation of this second-pass evaluation technique are shown in **Figure 10**.

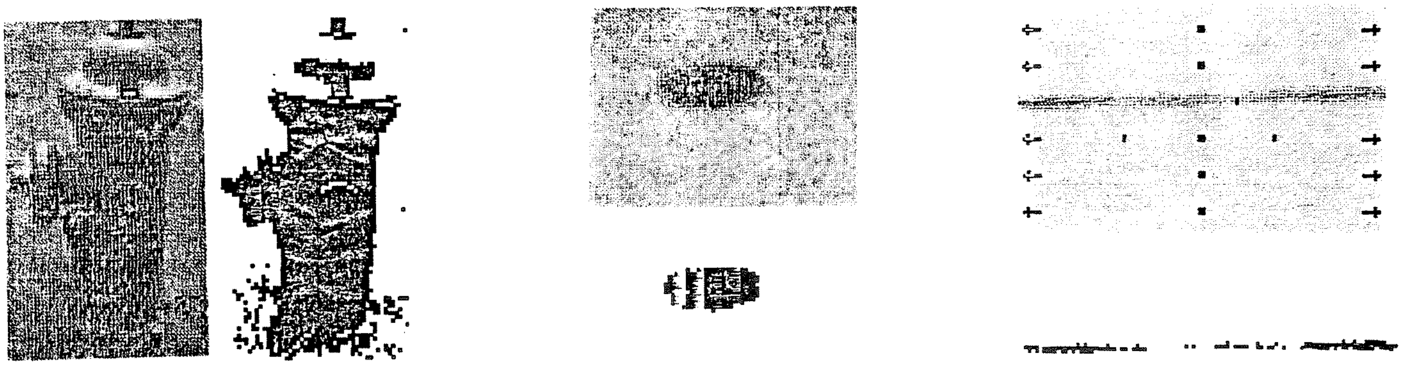
*Scene analysis for shadow identification*

In the 2-D image, a shadow is easily confused with objects, and this confusion degrades the performance of the vision system when it comes to the safety-path tracing. Also, from another perspective, shadow is recognized as an important feature for the interpretation of images (18,19). Therefore, it is important to be able to identify the presence of shadows in a scene.

The focus in the shadow identification approach is placed on the characterization of the inherent effect of a shadow, exploiting the fact that a shadow, when cast upon a given surface, preserves the intrinsic characteristics of the surface by virtue of the uniform effect of shadow. The characterization of the effect of shadow is supported, in this approach, by four interrelated analyses: 1) the histogram analysis; 2) the pixel intensity distribution (1-D intensity profiles) analysis; 3) the correlation analysis; and, 4) the power spectral analysis (20). The objective here is to analyze both the spatial domain and the frequency domain in order to determine through specific parameters that by going from the obstacle-free area to the assumed-shaded area, the surface physical characteristics have not changed but have only shifted by a uniform gray-level effect.

The procedure followed is to apply each analysis in a sequential fashion. If the first analysis does not yield a definitive answer, the second analysis will be conducted, and so on, until the last analysis is performed. In each analysis, performance parameters are established for shadow identification. If a shadow is positively identified at any stage, the shadow identification process terminates. Analysis in the frequency domain is left out as the last step due to its computational complexity.

The architecture of the proposed system for shadow identification is illustrated in **Figure 11**. The inputs to this



```
THE OBJECT MAY BELONG TO THE FOLLOWING CLASS:
[TREE-TRUNK, LIGHT POLE, FIRE HYDRANT, MAIL BOX, DOOR ENTRANCE...]
IT IS A LARGE SIZED OBJECT
IT IS IN FRONT OF YOU
IT IS 3.3 STEPS AWAY
```

```
THE OBJECT MAY BELONG TO THE FOLLOWING CLASS:
[SQUARE OR CIRCULAR SHAPED OBJECT]
IT IS A SMALL SIZED OBJECT
IT IS IN FRONT OF YOU
IT IS 4.0 STEPS AWAY
```

```
THERE IS NO OBJECT
FALSE ALARM BY THE FIRST-PASS EVALUATION
```

**Figure 10.**
Results of the second-pass evaluation.

system are selected subregions or windows of the image under consideration. The first of these windows is selected from the obstacle-free area and has been named window $W_F$. The second window is taken from the region containing the object and is called $W_O$. The third is selected so as to enclose partial segments of both regions and is called window $W_{FO}$ (**Figure 12**).

In outdoor scenes, shadows are of all shapes. Shadows that are cast by buildings, or large man-made objects, etc., have in general regular shapes and extend over a large area. These can be identified directly by the above approach. Unfortunately, we also have shadows that are cast by trees, sign posts, etc., which have irregular shapes. These shadows constitute a very difficult problem; however, there does appear to be a solution. Our approach involves a preprocessing step which would eliminate the gray-level effect, if uniform, in the assumed-shaded area. To make the assumed-shaded area look exactly like the obstacle-free area (if in fact the area in question is shaded), the thresholding technique of the second-pass evaluation is used to extract the assumed-shaded area.

To determine whether the eliminated gray-level effect is indeed that of a shadow, the procedure of shadow identification is reconducted. In this revision, the correlation results should improve. An implementation of this procedure on an outdoor scene is shown in **Figure 13**.

There are certain limitations to even the extended approach described. Unless additional information is provided, the following cases pose problems: 1) shadows cast on surfaces which have random texture or are marked by various irregularities in their intrinsic characteristics will disturb all the correlation measures; and, 2) dark shadows conceal all forms of intrinsic characteristics of the surface upon which they are cast.

Possible approaches to these problems are: (*a*) supplementation of the identification process with knowledge (e.g., shadows are not free-standing, their contours extend toward the object which cast them); and, (*b*) making use of additional information which can be provided by some electromagnetic device.

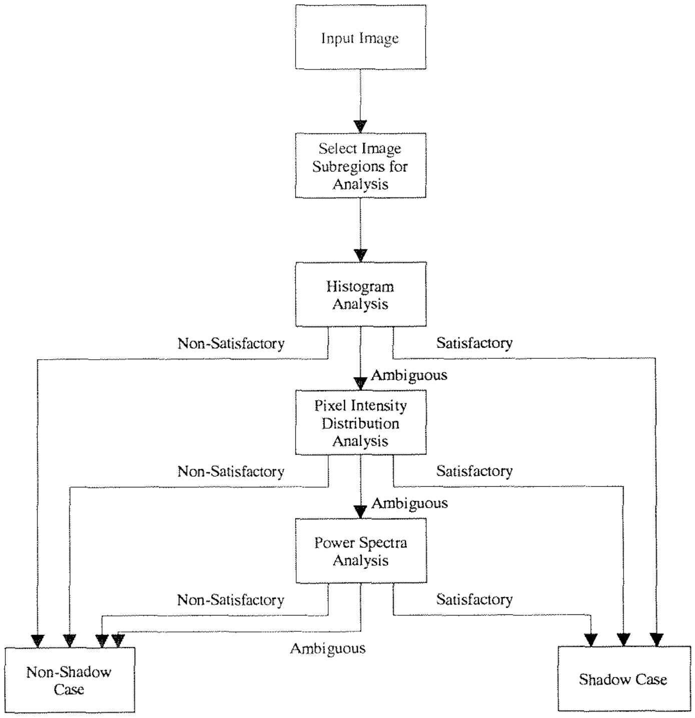


**Figure 11.**
Simplified architecture for shadow identification.

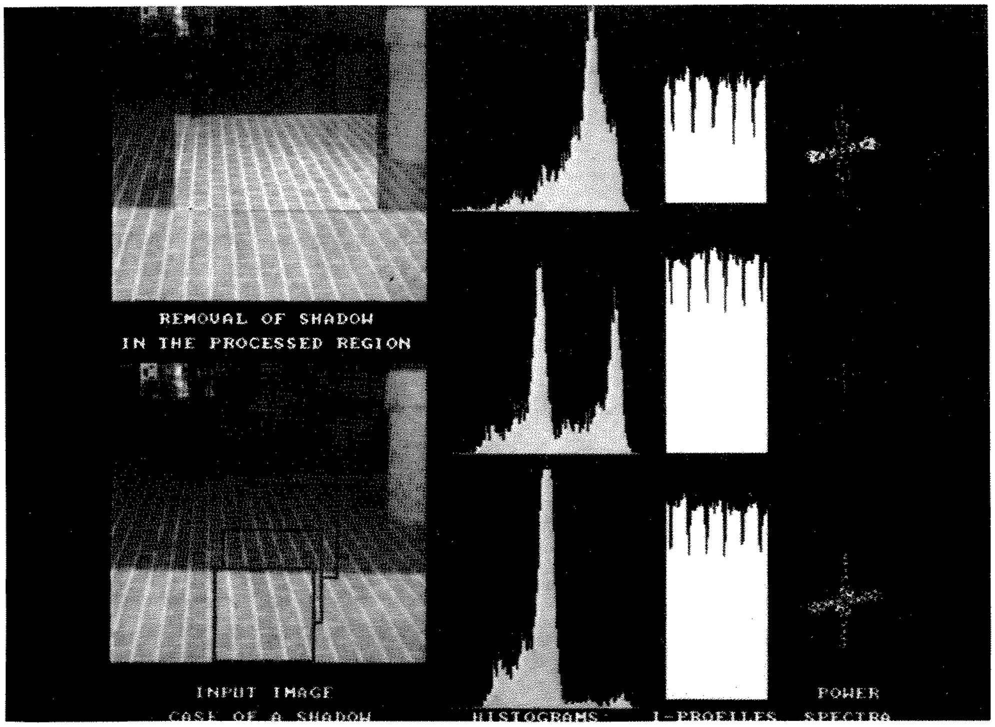


**Figure 12.**
Results of the shadow identification process.

*Scene analysis for the detection of depressions*

Depressions or drop-offs constitute a serious obstacle. Unfortunately, the detection of depressions is also a complex image analysis problem. In the human vision system, many visual cues (e.g., stereopsis, occlusion cues, context in the scene, and change in textural properties) are all integrated and interpreted with relative ease. In image processing, however, a computer implementation exploiting any one of the above cues becomes a complex information processing problem.

Clearly, there is no simple way to solve this problem. In this approach, we attempt to extract occluded information from a sequence of frames. This is based on the principle that if one is to approach a depression or a drop, one is bound to see new information which was previously occluded. This task, which necessitates analysis of a sequence of frames, requires image correspondence. The constraints of the image correspondence process are somewhat relaxed here since the concern is about locating, approximately, reference points in two different frames for the purpose of extracting occluded information (21). In this analysis, these reference points are chosen in the proximity where the obstacle is indicated by the first-pass evaluation. The procedure for extracting occluded information is as follows:

1. A specific window is set up on the vicinity of the detected obstacle.
2. Vertical and horizontal scans are taken to generate one-dimensional intensity profiles; these scans are delimitated by the size of the window.
3. Occluded information is checked for by comparing the major disturbances in the intensity profiles (both vertical and horizontal) from one frame to the next assuming a displacement toward the obstacle. A major disturbance is defined as any value exceeding the value $P_{max}$ given by

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

Parameters $\mu_p$ and $\sigma_p$ are the mean and standard deviation of the intensity profile, respectively.

Computer examples of this procedure are illustrated in **Figure 14**. When no occluded information is found in this analysis, the object remains a potential obstacle.

*Scene analysis of upright objects versus flat objects*

Distinguishing an upright object from a flat object is essential to a vision system. Upright objects may be obstacles to be avoided or landmarks which could help in the guidance process. Flat objects, on the other hand, could

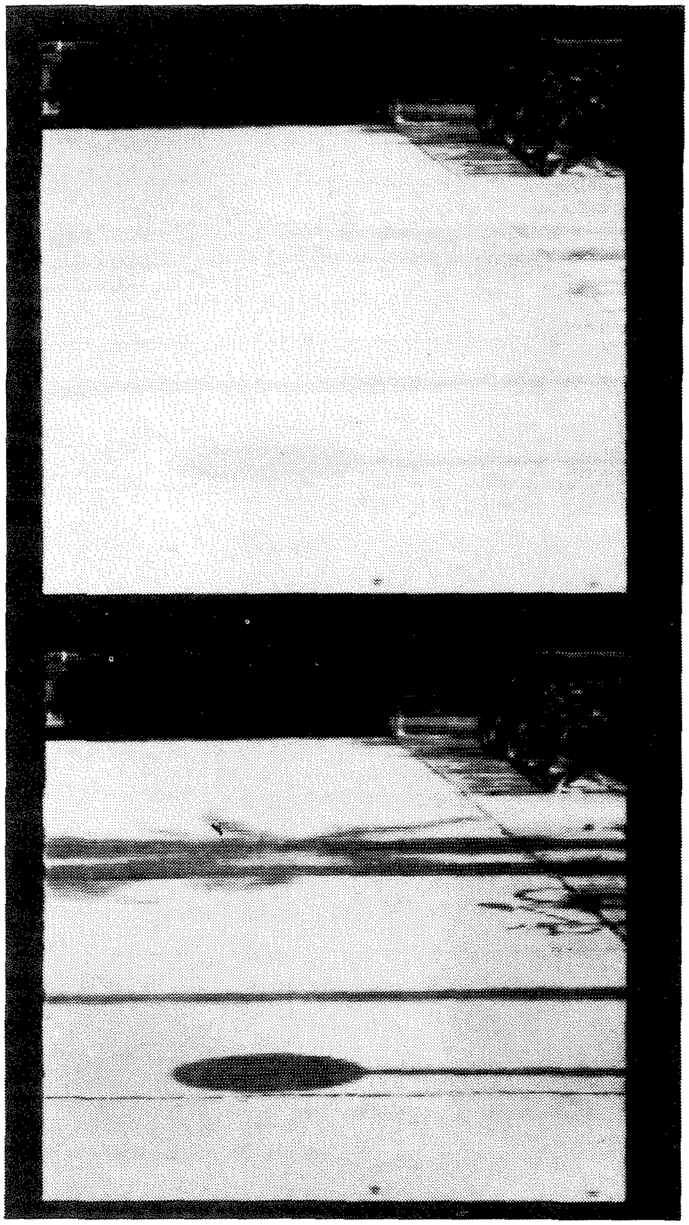


**Figure 13.**
Identification and removal of scattered outdoor shadows.

range from paper and other debris to texture change.

Some experimental observations were made contrasting the image projections of an upright object with that of a flat object. From these individual projections, the distinctive characteristics of the projections can be exploited to obtain a general technique to solve this problem.

*Observation 1.* Upright objects, unlike flat objects, are not affected by the perspective effect. Thus, for a fixed camera viewing position, objects with straight vertical edges will project as such on the 2-D image plane.

*Observation 2.* Upright objects project on the 2-D image plane proportionally to the extent (in length) of the area they occlude or the extent of information in the scene that is occluded.

*Observation 3.* Flat objects are affected by perspective. Also, flat objects project on the 2-D image plane proportional to their actual length (in the direction of travel) in terms of size of the object.

Technique to detect straight vertical edges

In this technique, Observation 1 is exploited and recourse made to edge detection. Since only the vertical edges of the image are of concern in this instance, a specialized edge-detection scheme was devised which makes use of the first derivative on the gray-level intensities. In the image, this derivative reduces to the difference in gray-level that exists between two adjacent pixels. This difference is used to evaluate the type of discontinuity in

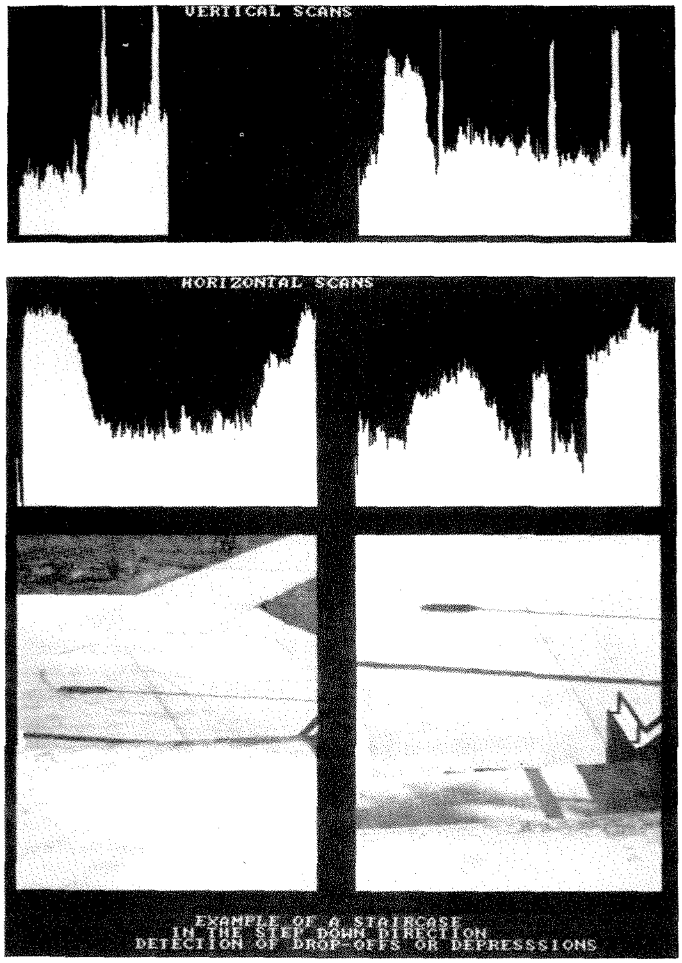
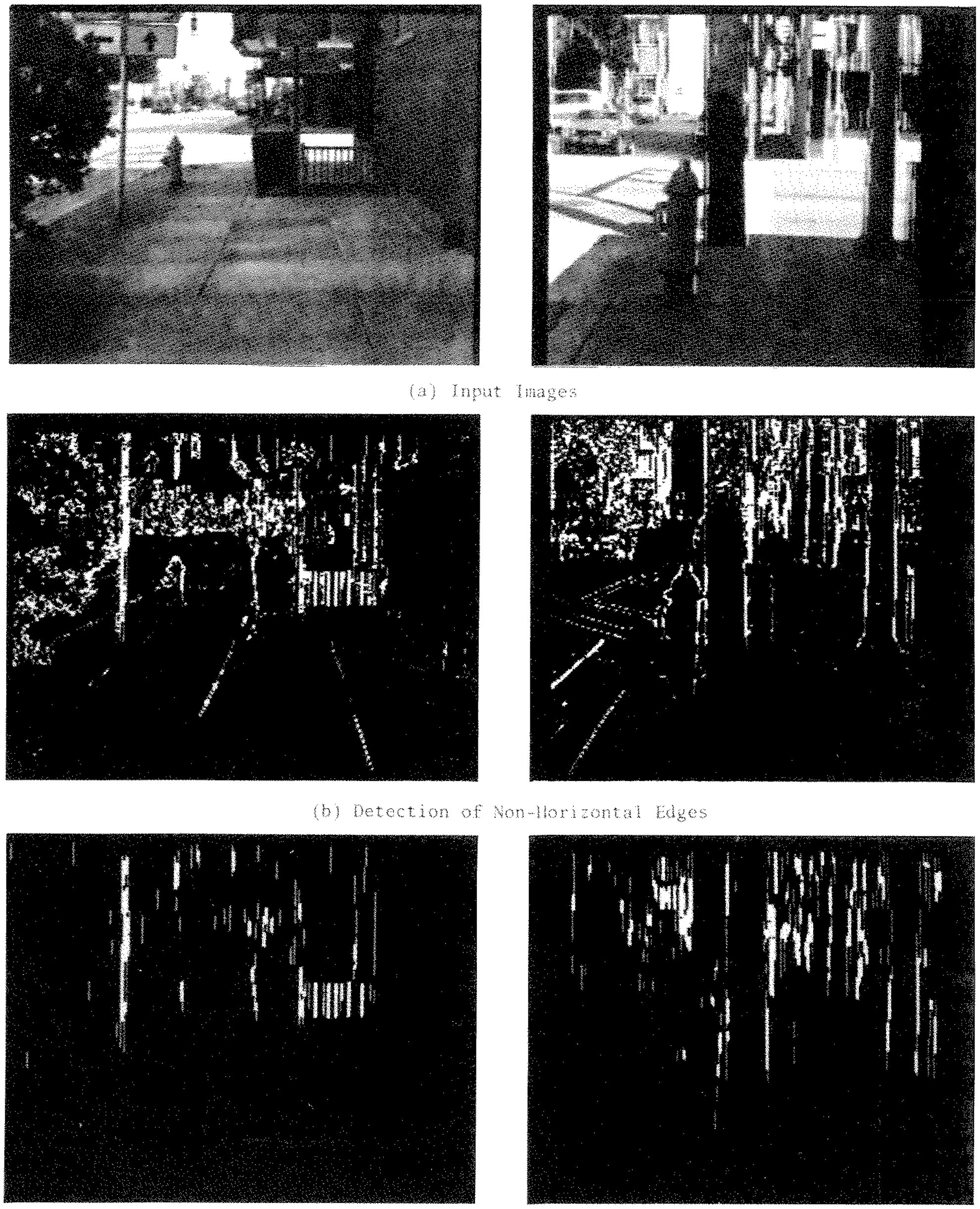


**Figure 14.**
Extraction of occluded information for the detection of depressions.



(a) Input Images

(b) Detection of Non-Horizontal Edges

(c) Detection of Vertical Edges Only

**Figure 15.**
Detection of upright objects.

intensity between the two pixels. A large discontinuity is a sign that an edge point may exist. Our approach to detect the vertical edges comprises two steps:

1. The first derivative is determined, pairwise, for all pixels in a horizontal scan of each line (record) of the image. Each time the derivative ($g_i - g_{i+1}$) exceeds a set threshold, $T_e$, point $g_{i+1}$ is considered a potential edge point. This process is repeated for all records of the image. To ensure that all potential edges are extracted, the threshold $T_e$ is set to a smaller value than $T_s$ as determined in the second-pass evaluation. Non-edge (noise) points are easily eliminated in the second step.
2. The system extracts those points with the same horizontal coordinate $x$ since the focus is on the vertical edges only.

The results of these two steps are illustrated in **Figure 15**. The locations of these upright objects are easily deter-

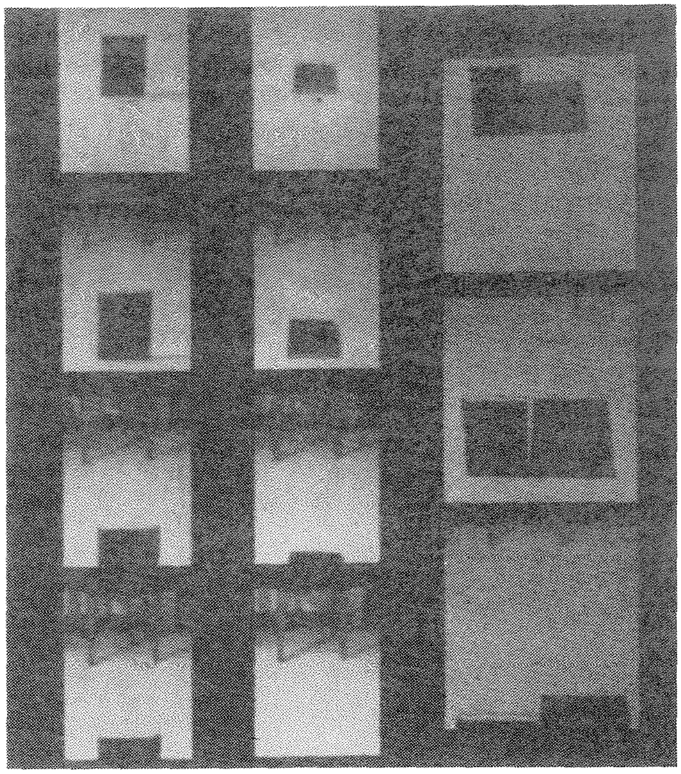


**Figure 16a.**
Moving toward upright versus flat objects.

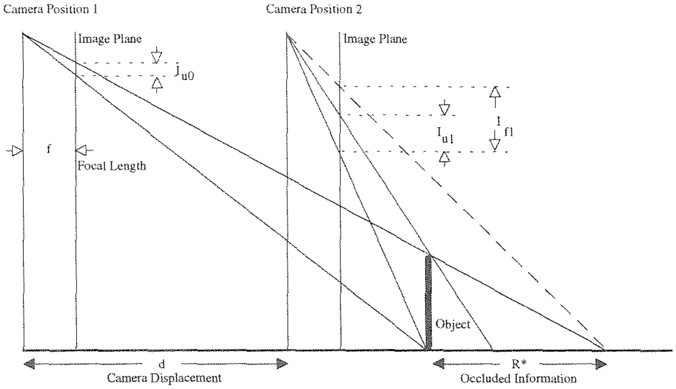


**Figure 16b.**
Analysis of upright versus flat objects: Geometric projections.

mined by their (*x,y*) coordinates. An edge-linking or contour-following technique can be used to enhance the results.

Picture projections of flat objects and upright objects

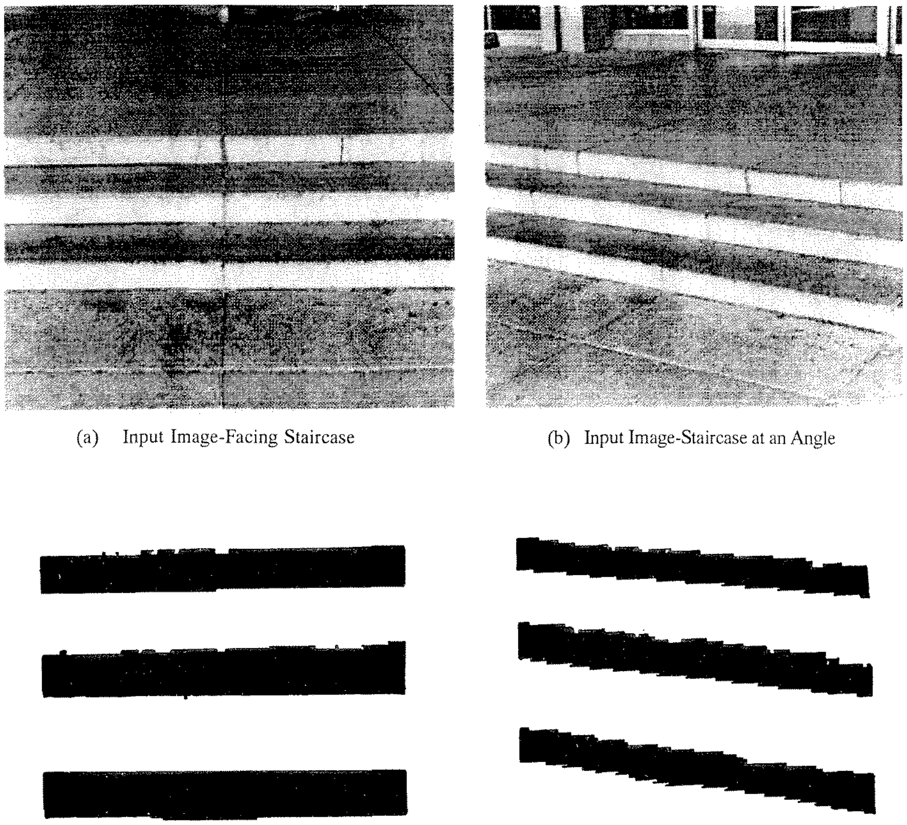
The measurement of these projections necessitates the generation of the binary image to allow for easy reference to the objects. Here, observations 2 and 3 are exploited. A solution to this complex analysis problem requires precise comparison of the resulting projections. It is found that the projection of a flat object on the image plane is proportional to its actual length. As a result, as an observer approaches a flat object there is an increase in the vertical projection of that object on the image plane; the increase is directly proportional to the actual distance moved by the observer. When this relationship holds for an object, we can deduce that this object is actually flat. In the case of an upright object, however, the picture plane projection is proportional to the length of the area it occludes, and this occluded length is a function of the range which separates object from observer (**Figure 16a**). These results can be verified using simple geometry. In **Figure 16b** we may assume, given $I_{uo}$, that the object is actually flat. In the subsequent frame, after a displacement, d, by the observer, we expect projection $I^*_{fl}$ but obtain $I_{ul}$ instead. This type of analysis can be carried out to verify whether an object is upright or flat.

Special case of the staircase example

The staircase is an interesting problem because, while shading generally distinguishes the riser (upright step) from the tread (flat step), the staircase is a succession of risers and treads. Therefore, to use the shadow identification process would result in an impression of flattening the staircase. For this reason, the first-pass evaluation is allowed to go on for as long as three obstacle warnings (i.e., to allow for 'ON' and 'OFF' type of path tracing); this is then used as an initial indication of the presence of a staircase (or perhaps a striped crosswalk). If the second-pass evaluation has also been used to provide a primitive description of the object, this primitive description can be used to enforce the notion that a staircase may indeed be present. With these primary cues, the recognition algorithm which identifies a staircase is initiated.

1. Obtain *m* vertical scans on the binary image ($3 \leq m \leq 5$) and determine the risers *r*(*s,k*) and treads *t*(*s,k*). Parameters *s* and *k* denote the step number of the staircase and the vertical scan line number, respectively.
2. Determine for all values *s* and *k* the ratio
$$R(s,k) = \frac{r(s,k)t(s,k)}{r(s,k) + t(s,k)}$$
this ratio is used in agreement with the standard set by the building codes.
3. If this ratio has about the same value for all *m* vertical scans, a staircase is identified.

For striped crosswalks (if one desires to include them in the identification), steps 1, 2, and 3 above can be repeated



(a) Input Image-Facing Staircase

(b) Input Image-Staircase at an Angle

(c) Window of Binary, Noise-Free Gap-filled Image of (a)

(d) Window of Binary, Noise-Free Gap-Filled Image of (b)

IDENTIFICATION RESULTS

THE OBJECT IS A STAIRCASE
YOU ARE FACING THE OBJECT

IDENTIFICATION RESULTS

THE OBJECT IS A STAIRCASE
THE OBJECT IS AT YOUR LEFT

**Figure 17.**
Results of the staircase identification process.

except that the ratio $R(s,k)$ for a crosswalk is larger than the ratio $R(s,k)$ for a staircase.

To include orientation, with respect to the observer (user) the ratio $\alpha_r = r(s,1)/r(s,m)$ can be used to identify the following situations:

$\alpha_r < 1$ the staircase is to the left
$\alpha_r = 1$ the staircase is straight ahead
$\alpha_r > 1$ the staircase is to the right, and the upright versus the flat object analysis would simply indicate that the object is an upright object to the right.

A computer implementation of this technique is shown in **Figure 17**. With new ideas, such as those proposed by Sakamoto and Mehr (22), the staircase problem will be even easier to solve.

## SUMMARY

In the effort to develop a man-machine vision interface as an environment sensing device for individuals with a visual impairment, the following problems were identified and addressed: 1) the vision system design; 2) establishment of the mapping principles between the 2-D images and the 3-D real world; 3) development of imaging techniques; and, 4) establishment of the appropriate communication link between the vision system and the visually impaired individuals.

The research efforts have thus far yielded algorithms which, under certain constraints, recover the depth information from 2-D images; and imaging techniques which (*a*) plan a safety path and provide guidance cues, (*b*) detect drop-offs or depression, (*c*) discriminate upright objects from flat objects, (*d*) identify shadows (false alarms), and, (*e*) identify important objects such as stairs and crosswalks.

The imaging techniques studied, aside from those that deal with the frequency domain, can be implemented in real-time once parallel processing schemes become an integral part of the vision system. Most of the processing time at present is spent reading the input image and writing the output image from and into memory. The output image is created simply for viewing purposes. In future implementations, the input images should be read directly from the sensing arrays of the camera(s). Those imaging techniques which do not lend themselves to real-time processing should be implemented using application-specific integrated circuits (ASICs). Parallel processing should be a criterion to be considered at all levels, from the devising of the imaging techniques to the building of the structure which will integrate and govern these imaging techniques.

When such a system might be put to practical application depends upon substantial research and development efforts by all concerned with the mobility problem. The challenging technical issues must be matched with adequate economic justification for the costly research and development efforts, beyond which are the issues of manufacturing and marketing a future device. The market for electronic travel aids remains difficult to define and the visually impaired public thus far seems unimpressed with extant approaches and persists in relying upon a sighted companion, a dog, or a long stick. On a more positive note, the computer vision approach to the mobility of visually impaired persons capitalizes on the substantial and ongoing investment in research, development, and application of computer vision. Computer vision is a research area in its own right with adherents in the artificial intelligence community, in neural and cognitive science, in manufacturing, and in military applications. "Spin-offs" from these more affluently supported arenas can make important contributions to the application of computer vision to safe and effective travel for persons with visual impairments. Still, a great deal of research and implementation work remains to be done before such a machine vision interface can be put to practical use beyond the laboratory and into the real world.

## ACKNOWLEDGMENT

The author wishes to express his gratitude for the support provided by the Department of Veterans Affairs and the Pacific International Center for High Technology Research in Honolulu, Hawaii.

## REFERENCES

1. Zahl PA, editor. Blindness: modern approaches to the unseen environment. Princeton: Princeton University Press, 1950.
2. Veterans Administration Rehabilitation R&D Progress Reports, 1985–present.
3. Committee on Vision, National Research Council. Electronic travel aids: new directions for research. Washington, DC: National Academy Press, 1986.
4. Brabyn JA. New development in mobility and orientation aids for the blind. IEEE Trans Biomed Eng 1982 Apr; BME-29(4):285-9.
5. Blasch BB, Long RG, Griffin-Shirley N. Results of a national survey of electronic travel aid use. J Vis Impairm Blindn 1989 Nov;83(9):449-53.
6. Collins CC, Saunders FA. Pictorial display by direct electrical stimulation of the skin. J Biomed Syst 1976;1(2):3-16.
7. Loomis JM. Tactile pattern perception. Perception 1981;10:5-27.
8. Komoriya K, Tachi S, Tanie K, Ohno T, Abe M. A method for guiding a mobile robot using discretely placed landmarks. J Mech Eng Lab 1983;37(1):1-10.
9. Deering MF. Real time natural scene analysis for a blind prosthesis. Mountain View, (CA): Fairchild Corporation: 1982 Aug. Technical Report, No. 622.
10. Thorpe CE. FIDO: Vision and navigation for a robot rover [dissertation]. Pittsburgh, (PA): Carnegie Mellon Univ. 1986.
11. Moravec HP. Pittsburgh (PA): Carnegie Mellon University. Robotics Inst.: 1980 Sept. Technical Report No. CMU-RI-TR3.
12. Gennery DB. A stereo vision system for an autonomous vehicle. In: Proceedings of the 5th International Joint Conference on Artificial Intelligence, 1977; Cambridge, MA: MIT Press: 576-80.
13. Inigo RM, McVey ES, Berger BJ, Wirtz MJ. Machine vision applied to vehicle guidance. IEEE Trans Pattern Anal Mach Intel 1984 Nov;6(6):820-6.
14. Marr D. Vision. Cambridge, MA: MIT Press, 1979.
15. Adjouadi M, Zhang XB. Stereo matching analysis. Proceedings of the SouthCon 92 Conference, 1992 March 10-12; Orlando, FL.
16. Negahdaripour S, Horn BKP. Direct passive navigation. IEEE Trans 1987 Jan;PAMI-9(1):168-76.
17. Horn BKP, Weldon EJ Jr. Robust direct methods for recovering motion. Int J Comput Vision 1988;2:51-76.
18. Shafer SA, Kanade T. Using shadows in finding surface orientations. Comput Vision Graphics Image Process 1983 Apr;22:145-76.
19. Horn BKP. Obtaining shape from shading information. In: Horn BKP, Brooks MJ, editors. Shape from shading. Cambridge, MA: MIT Press, 1989: 123-71.
20. Tou JT, Adjouadi M. Shadow analysis in scene interpretation. In: Proceedings of the 4th Scandinavian Conference on Image Analysis, June 1985, Trondheim, Norway.
21. Adjouadi M. Image techniques for the detection of depressions in autonomous guidance. Vision Interface '86, Vancouver, BC, Canada, May 1986.
22. Sakamoto L, Mehr EB. A new method of stair markings for visually impaired people. J Visual Impairm Blindn 1988 Jan;82(1):24-7.

## APPENDIX A: WIDE-ANGLE VIEW TECHNIQUE

### a. Configuration of the Wide-Angle View

A configuration of the wide-angle view is illustrated in **Figure A.1(a)**. This configuration is characterized by two angles of view of the camera. These are the horizontal angle of view and the vertical angle of view. From simple triangulation, these two angles are determined using the relation below:

$$v(q)=2\arctan(\frac{q}{2f})$$

where q depends on the type of film used. For example, if we use a 35 mm lens (f = 35 mm) and a 35 mm film whose single-frame dimension is 35x24 mm, then, using the above equation, the horizontal angle of view, $\theta$, is derived by substituting the frame length for q, as $\theta = V(35) \approx 53°$, and the vertical angle of view, $\phi$, is derived by substituting the frame width for q, as $\phi = V(24) \approx 35°$.

### b. Integration Process of the Wide-Angle View

Each image of the wide-angle view is processed using the first-pass evaluation technique, and all the necessary parameters of the wide-angle view have been determined. With reference to **Figure A.1(b)**, parameters Pl, Pf, and Pr correspond to the safety paths for left, front, and right image. Parameters $C_{ij}$ denote the various clearances where the first index, i, identifies the image and has values of l, f, and r for left, front and right images, respectively. The second index, j, identifies the direction and has values of l and r for left and right direction respectively. Given these parameters, the integration process which yields an optimal safety path is performed using the step-wise procedure described below:

(1) Determine the left and right optimal clearances $C_l$ and $C_r$ of the wide-angle view given by

$$C_l=\min(C_{fl},C_{lr})$$
$$C_r=\min(C_{fr},C_{rl})$$

(2) If clearance $C_l$ and the path of the left image $P_l$ exist, determine both the angle and path of the wide-angle view in the left direction.

(a) The angle of the wide-angle view in the left direction is given by

$$\lambda_l=\arctan[\frac{\min(C_l,P_l)}{L_w}]$$

where $L_w$ is a function of the horizontal angle of view. For example, if 53° as derived earlier, then $L_w \approx L$. L is, as defined earlier, the range between the camera and the nearest point viewed by the camera.

(b) The path of the wide-angle view in the left direction is given by

$$P_{\lambda l}=\frac{L_w}{\cos(\lambda_l)}$$

(3) Perform a similar analysis (as in step 2) for the wide-angle view in the right direction.

(a) The angle of the wide-angle view in the right direction is given by

$$\lambda_r=\arctan[\frac{\min(C_r,P_r)}{L_w}]$$

**Figure A.1.**
The wide-angle view technique.

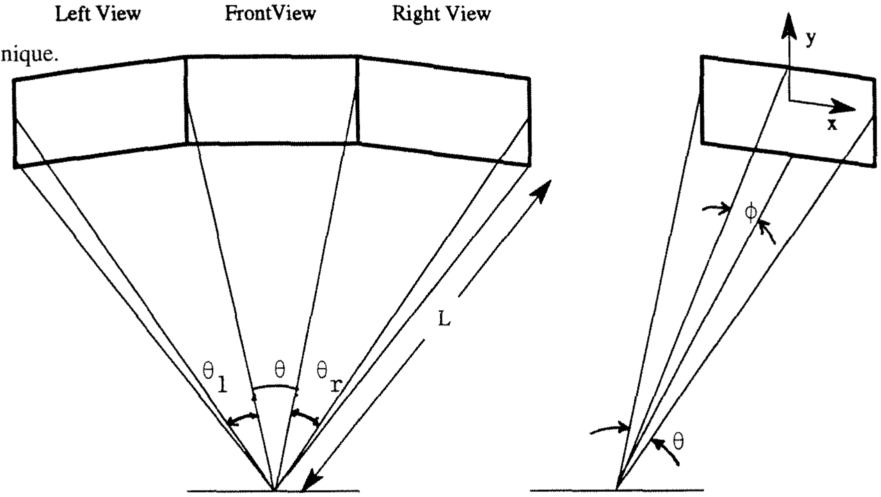


**Figure A.1a.**
Configuration of the wide-angle view.

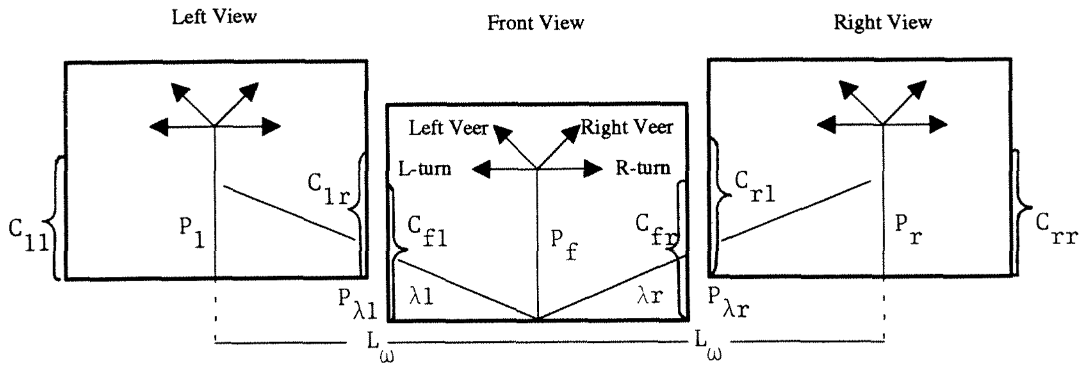


**Figure A.1b.**
Parameters of the wide-angle view.

(b) The path of the wide-angle view in the right direction is given by

$$P_{\lambda r}=\frac{L_w}{\cos(\lambda_r)}$$

(4) In the case where the wide-angle view in the left direction is chosen, determine the remaining portion of path $P_l$ denoted by $P_{ll}$ which extends beyond path $P_{\lambda l}$, and is given by

$$P_{ll}=P_l-L_w\tan(\lambda_l)$$

(5) Similarly, in the case where the wide-angle view in the right direction is chosen, determine the remaining portion of path $P_r$ denoted by $P_{rl}$ which extends beyond path $P_{\lambda r}$, which is given by

$$P_{rl}=P_r-L_w\tan(\lambda_r)$$

(6) The additional information of veers or turns which may take place after either path $P_l$ and $P_r$, can be determined using the same analysis as that performed in steps 2(a) and 2(b) or 3(a) and 3(b), respectively.

## APPENDIX B: PROCEDURE OF THE SECOND-PASS EVALUATION TECHNIQUE

(1) Quantize the average gray level values of $G(C_k)$ as follows:

$$\begin{gathered}
G(C_k) \leq 10 \rightarrow C_k=0 \\
10 < G(C_k) \leq 35 \rightarrow C_k=0.25 \\
35 < G(C_k) \leq 55 \rightarrow C_k=0.50 \\
55 < G(C_k) \leq 75 \rightarrow C_k=0.75 \\
G(C_k) \leq 75 \rightarrow C_k=1
\end{gathered}$$

(2) Determine the cumulative cell values given b

(a) for a horizontal overlay

$$H_c(l_1)=\sum_{i=1}^{n} C_{i+n(l-1)} \qquad l_1=1,2,\ldots,n$$

(b) for a vertical overlay

$$V_c(k_1)=\sum_{i=1}^{n} C_{k+n(i-1)} \qquad k_l=1,2,\ldots,n$$

(c) for a diagonal overlay with a left inclination

$$D_{rl}(k_l)=\sum_{i=1}^{k_1} C_{k_1+(i-1)(n-1)(n-k)} \qquad k_1=1,2,\ldots,n$$

and

$$D_{r2}(k_1)=\sum_{i=1}^{k_1} C_{k_1+(i-1)(n-1)} \qquad k_1=1,2,\ldots,n-1$$

(d) for a diagonal overlay with a right inclination

$$D_{l1}(k_1)=\sum_{i=1}^{k_1} C_{(n+1)i+(n-k_1)n} \qquad k_l=1,2,\ldots,n$$

and

$$D_{l2}(k_l)=\sum_{i=1}^{k_l} C_{(n+1)i-k_1} \qquad k_l=1,2,\ldots,n-1$$

With the above steps, the following assessment about the object is made.

(1) If all values of $H_c(1_1)=0$ $1_1=1,2,\ldots,n$, this indicates a false alarm by the first-pass evaluation (assuming reasonable range). This can be the case of debris dirt spots on the path of travel.

(2) Determine the general overlay of the object by finding

$$K_{max}=\max[H_c,V_c,D_{r1},D_{r2},D_{l1},D_{l2}]$$

for example, if $K_{max}=H_c$, then the object has a horizontal overlay.

(3) Categorize the object using the aspect ratio below

$$A_r=\frac{\max[V_c,D_{r1},D_{r2},D_{l1},D_{l2}]}{\max[H_c]}$$

If $A_r > 1.25$: tree trunk, fire hydrant, light pole, etc.
If $0.75 \leq A_r \leq 1.25$: square or circular shaped object
If $A_r < 0.75$: curb, step, bench, etc.

(4) Determine a more accurate range and location of the object with respect to the user using a point in the object whose coordinates ($x_p$, $y_p$) are such that $y_p$ is the nearest point with respect to the user, and $x_p$ is a point nearest to the center of the path of travel.

(5) The approximate size of the object is estimated using the following relation

$$S_{obj}=(D_{obj})[\max(H_c)]$$

where

$$D_{obj}=\max(V_c,D_{r1},D_{r2},D_{l1},D_{l2}).$$