BOOTSTRAP STEREO

Marsha Jo Hannah

Lockheed Palo Alto Research Laboratory
Department 52-53, Building 204
3251 Hanover Street, Palo Alto, CA 94304

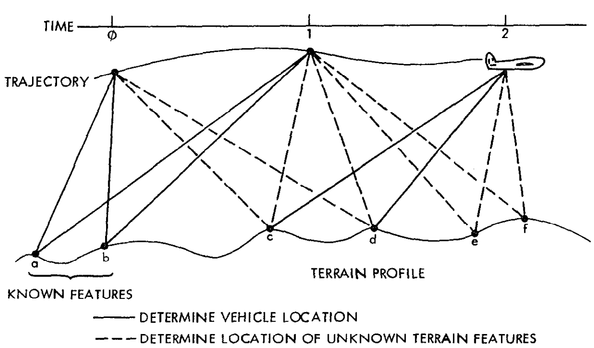
ABSTRACT

Lockheed has been working on techniques for navigation of an autonomous aerial vehicle using passively sensed images. One technique which shows promise is bootstrap stereo, in which the vehicle's position is determined from the perceived locations of known ground control points, then two known vehicle camera positions are used to locate corresponding image points on the ground, creating new control points. This paper describes the components of bootstrap stereo.

I INTRODUCTION

Before the advent of sophisticated navigation aids such as radio beacons, barnstorming pilots relied primarily on visual navigation. A pilot would look out the window of his airplane, see landmarks below him, and know where he was. He would watch the ground passing beneath him and estimate how fast and in what direction he was moving.

Today, there exist applications for which a computer implementation of this simple, visually oriented form of navigation would be useful. One scenario hypothesizes a small, unmanned vehicle which must fly accurately from its launch point to its target under possibly hostile circumstances.



TIME	POSITION OF CRAFT DERIVED FROM	DETERMINE POSITION OF
0	a, b	–
1	a, b	c, d
2	c, d	e, f

Figure 1 Navigation Using Bootstrap Stereo.

Our overall approach to the problem involves providing the vehicle with a Navigation Expert having approximately the sophistication of an early barnstorming pilot. This expert will navigate partly by its simple instruments (altimeter, airspeed indicator, and attitude gyros), but mostly by what it sees of the terrain below it. This paper covers one aspect of the Navigation Expert, a technique which we call bootstrap stereo.

II THE BOOTSTRAP STEREO CONCEPT

Given a set of ground control points with known real-world positions, and given the locations of the projections of these points onto the image plane, it is possible to determine the position and orientation of the camera which collected the image. Conversely, given the positions and orientations of two cameras and the locations of corresponding point-pairs in the two image planes, the real-world locations of the viewed ground points can be determined [1]. Combining these two techniques iteratively produces the basis for bootstrap stereo.

Figure 1 shows an Autonomous Aerial Vehicle (AAV) which has obtained images at three points in its trajectory. The bootstrap stereo process begins with a set of landmark points, simplified here to two points a and b, whose real-world coordinates are known. From these, the camera position and orientation are determined for the image frame taken at Time 0. Standard image-matching correlation techniques [2] are then used to locate these same points in the second, overlapping frame taken at Time 1. This permits the second camera position and orientation to be determined.

Because the aircraft will soon be out of sight of the known landmarks, new landmark points must be established whenever possible. For this purpose, "interesting points" -- points with a high likelihood of being matched [3] -- are selected in the first image and matched inthe second image. Successfully matched points have their real-world locations calculated from the camera position and orientation data, then join the landmarks list. In Figure 1, landmarks c and d are located in this manner at Time 1; these new points are later used to position the aircraft at Time 2. Similarly, at Time 2, new landmarks e and f join the list; old landmarks a and b, which are no longer in the field of view, are dropped from the landmarks list.

Once initialized from a set of known landmarks, bootstrap stereo has four components -- camera calibration, new landmark selection, point matching, and control point positioning. Because camera calibration and control point positioning have been well covered in the photogrammetric and imaging literatures (e.g., [1], [4], [5], [6]), we will discuss only landmark selection and point matching in the following sections.

III NEW LANDMARK SELECTION

Because the aircraft rapidly moves beyond the known landmarks, new landmark points must constantly be established. For this purpose, "interesting points" -- points with a high likelihood of being matched [3] -- are selected in the old image of each pair, then matched with their corresponding points in the new image and located on the terrain.

Matching is done on the basis of the normalized cross-correlation between small windows of data (typically 11×11) around the two points in question. Matching has trouble in areas that contain little information or whose only information results from a strong linear edge, therefore such areas make poor candidate landmarks.

To avoid mismatches from attempting to use such aras, various measures on the information in the window have been used, including the simple statistical variance of the image intensities over the window [2] and the minimum of the directed variances over the window [3]. We have combined these into another interest measure which we call edged variance, which appears to perform better than either of its components [7].

We have defined our interesting points to be those which are local peaks in our interest measure, with a lower bound established to reject undesirable areas. Figure 2 includes some examples of the application of this interest measure.

IV POINT MATCHING

The actual matching of points in an image pair is done by maximizing normalized cross-correlation over small windows surrounding the points. Given an approximation to the displacement which describes the match, a simple spiraling grid search is a fairly efficient way to refine the precise match [2]. To provide that initial approximation, we have employed a form of reduction matching [3].

We first create a hierarchy of N-ary reduction images. Each $N \times N$ square of pixels in an image is averaged to form a single pixel at the next level. This reduction process is repeated at each level, stopping when the image becomes approximately the size of the correlation windows being used. Matching then begins at the smallest images, with the center point of the first image being matched via a spiral search. Thereafter, each matched point spawns four points around itself, offset by half a window radius along the diagonals of the window. These are mapped down to the next level of images, carrying their parent's displacement (suitably magnified) as their suggested match approximation. These matches are refined by a spiraling search before spawning new points. This process continues until the largest images are reached, effectively setting up a grid of matched points.

In our implementation of bootstrap stereo, reduction matching is used to determine approximate registration of the images and to initialize the second-order match prediction polynomials. Matching of old landmarks and of interesting points to create new landmarks uses these polynomials to predict an approximate match, which is then refined by a local search. Autocorrelation thresholding is used to test the reliability of the match, then points are located more closely than the image grid permits by parabolic interpolation of the X- and Y-slices of the correlation values.

V AN EXAMPLE

In Figure 2, we present an example of the control-point handling portion of bootstrap stereo. The original data set, a sequence of 3 images from a video tape taken over the Night Vision Laboratory terrain model, is shown in Figure 2a.

Figure 2b shows the interesting points in the first image, indicated by + overlays. If these were the control points from a landmark processor, we would use them to locate the first camera. These landmark points are next matched with their corresponding points in the second image; Figure 2c shows the successful matches overlaid on the first and second images. From the image plane positions of these points, the position and orientation of the second camera are determined.

Next, the areas of the second image which were not covered by matches are blocked out and interesting points are found in the uncovered areas, as seen in Figure 2d. The old landmark points and the interesting points are then matched in the third image, as shown in Figure 2e. The old control points from the second image are used to calibrate the third camera; the camera calibrations are then used to locate the matched interesting points on the ground, forming new control points. These two steps are then repeated for subsequent pairs of images in longer sequences.

VI CONCLUSIONS

When an autonomous aerial vehicle must navigate without using external signals or radiating energy, a visual navigator is an enticing possibility. We have proposed a Navigation Expert capable of emulating the behavior of an early barnstorming pilot in using terrain imagery. One tool such a Navigation Expert could use is bootstrap stereo. This is a technique by which the vehicle's position is determined from the perceived positions of known landmarks, then uses two known camera positions to locate real-world points which serve as new landmarks.

The components of bootstrap stereo are well established in the photogrammetry and image processing literature. We have combined these, with improvement, into a workable system. We are working on an error simulation, to determine how the errors propagate and accumulate.

VII REFERENCES

[1] Thompson, M. M., Manual of Photogrammetry, American Society of Photogrammetry, Falls Church, Virginia, 1944.

[2] Hannah, M. J., Computer Matching of Areas in Stereo Imagery, PhD Thesis, AIM#239, Computer Science Department, Stanford University, California, 1974.

[3] Moravec, H. P., "Visual Mapping by a Robot Rover", Proceedings of the 6th IJCAI, Tokyo, Japan, 1979.

[4] Duda, R. O. and P. E. Hart, Pattern Classification and Scene Analysis, John Wiley and Sons, New York, New York, 1973.

[5] Fischler, M. A. and R. C. Bolles, "Random Sampling Consensus", Proceedings: Image Understanding Workshop, College Park, Maryland, April 30, 1980.

[6] Gennery, D. B., "A Stereo Vision System for an Autonomous Vehicle", Proceedings of the 5th IJCAI, Cambridge, Massachusetts, 1977.

[7] Hannah, M. J., "Bootstrap Stereo", Proceedings: Image Understanding Workshop, College Park, Maryland, April 30, 1980.

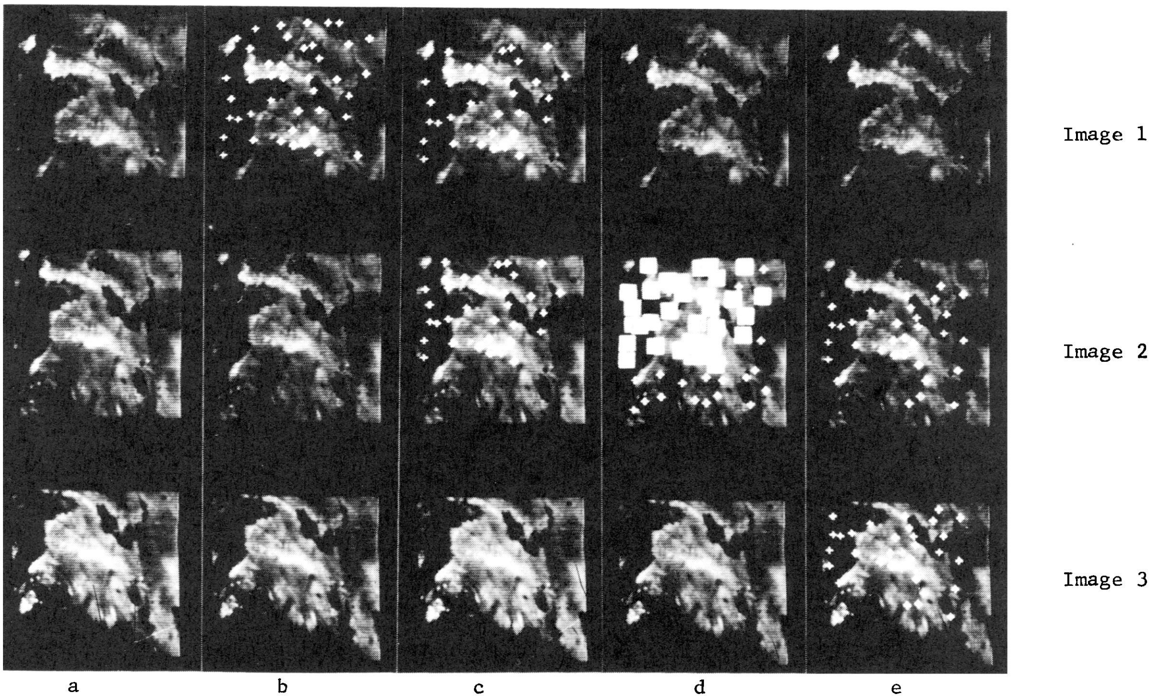


Figure 2 An Example of the Control-Point Handling for Bootstrap Stereo

a) The original sequence of 3 images.

b) The interesting points in Image 1.

c) The matched points between Images 1 and 2.

d) The areas of Image 2 covered by matches, with interesting points found in the uncovered areas.

e) The control points in Image 2 matched to Image 3.