REPORT NO. DOT-TSC-FHWA-75-1

AUTOMATIC EXTRACTION OF HIGHWAY TRAFFIC DATA FROM AERIAL PHOTOGRAPHS

Juris G. Raudseps

APRIL 1975

FINAL REPORT

DOCUMENT IS AVAILABLE TO THE PUBLIC THROUGH THE NATIONAL TECHNICAL INFORMATION SERVICE, SPRINGFIELD, VIRGINIA 22161

Prepared for

U.S. DEPARTMENT OF TRANSPORTATION

FEDERAL HIGHWAY ADMINISTRATION
Associate Administrator for Research and Development
Office of Research
Washington DC 20590

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT-TSC-FHWA-75-1		

4. Title and Subtitle

AUTOMATIC EXTRACTION OF HIGHWAY TRAFFIC DATA FROM AERIAL PHOTOGRAPHS

5. Report Date: April 1975

6. Performing Organization Code:

7. Author(s)

Juris G. Raudseps

8. Performing Organization Report No.: DOT-TSC-FHWA-75-1

9. Performing Organization Name and Address

U.S. Department of Transportation
Transportation Systems Center
Kendall Square
Cambridge MA 02142

10. Work Unit No.: FCP 31C3 532

11. Contract or Grant No.: HW405/R4205

12. Sponsoring Agency Name and Address

U.S. Department of Transportation
Federal Highway Administration
Associate Administrator for Research and Development, Office of Research
Washington DC 20590

13. Type of Report and Period Covered: Final Report July 1970-June 1974

14. Sponsoring Agency Code:

15. Supplementary Notes

16. Abstract

The design of a system for scanning sequences of aerial photographs with a computer-controlled flying-spot scanner and automatically measuring vehicle locations is described. Hardware and software requirements for an operational system of this type are enumerated. Measurement accuracy is predicted to be comparable to that achieved with manual methods in high-volume applications. The cost of such a system is estimated to exceed $500,000. Efficient operation is shown to be critically dependent on the development of an algorithm for predicting vehicle positions that is significantly better than that now available.

17. Key Words	18. Distribution Statement
Aerial Photography Highway Traffic Pattern Recognition	DOCUMENT IS AVAILABLE TO THE PUBLIC THROUGH THE NATIONAL TECHNICAL INFORMATION SERVICE, SPRINGFIELD, VIRGINIA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	78	

Form DOT F 1700.7 (8-69)

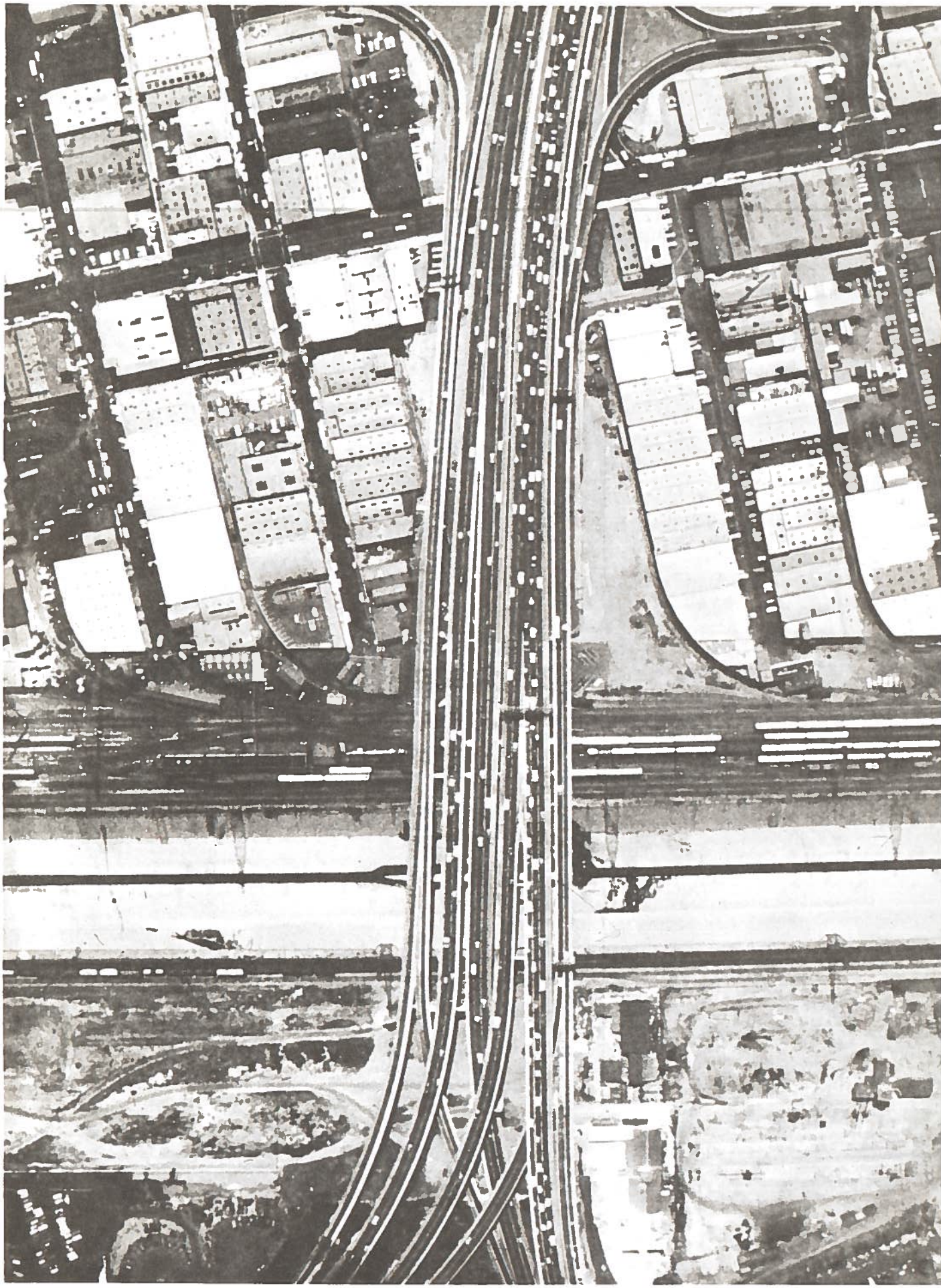
PREFACE

The work reported here was performed at the Transportation Systems Center (TSC) for the Office of Research and Development of the Federal Highway Administration. The project monitor was Harry S. Lum, HRS 33. The objective of this effort was to investigate the applicability of advanced image processing techniques to the problem of automatically extracting traffic data from aerial photographs. Aerial photographs are useful in various studies of highway traffic behavior, but their utilization involves time consuming and expensive data reduction. A system for sharing this task between man and automatic film-scanning equipment was designed and evaluated.

A computer-controlled flying-spot scanner system was assembled at TSC to serve as an experimental facility, and numerous machine-language computer programs were written to operate it. Various applicable program listings are given in an interim report (DOT-TSC-FHWA-71-1) by the current author and Dr. David S. Prerau. Richard H. Robichaud, Lennart E. Long, and William R. Murphy at TSC contributed significantly in overcoming the numerous difficulties with the equipment to which this undertaking was subject.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS



1. INTRODUCTION

This report describes the work performed at the DOT Transportation Systems Center in Cambridge, Massachusetts for the Federal Highway Administration in investigating techniques and the overall feasibility of automating the the process of extracting highway traffic information from aerial photographs recording such traffic.

1.1 APPLICATIONS OF AERIAL PHOTOGRAPHS OF TRAFFIC

There is considerable interest in gathering data on traffic flow by means of aerial photography. The major reason for this is that an aerial photograph is the only practical medium for observing the location of every vehicle within the area covered at the same given instant in time. By extension, in successive frames of photography covering a given area one can observe all the changes in vehicle locations that have occurred during the interval between the instants that the photographs were taken. From these, average vehicle velocities over the interval can be calculated.

Both macroscopic and microscopic aspects of traffic flow can be observed from aerial photographs. The term "macroscopic" in this context denotes features pertaining to the flow of traffic as a whole - such things as mean velocity, density, flow rate in vehicles/hour, distribution of traffic by lanes, by vehicle types (cars, trucks), etc. "Microscopic" denotes those features defined in terms of the actions of individual vehicles or configurations of vehicles. Included among them are such things as headways between individual vehicles, lane-changing behavior, etc. Generally speaking, those concerned with highway operation would be interested in the macroscopic aspects of flow - in such operationally meaningful quantities as the number of vehicles on a given stretch of road, the average velocity along that stretch, travel time between some pair of points, etc. The microscopic aspects of traffic are more of concern to designers, who must be

able to predict the effects upon traffic flow of ramps, curves and grades, etc., and theoreticians, who need data on which to base their models and with which to validate them.

Sequences of aerial photographs inherently contain a microscopic depiction of traffic behavior over the period they cover. This property makes aerial photography a practically irreplaceable tool in those situations where the precise behavior of individual vehicles is the prime matter of interest, since the amount of instrumentation and recording equipment required to follow simultaneously and precisely a large number of vehicles over an extended area is in general too formidable to consider.

It is less obvious that aerial photography is a good tool for observing the macroscopic aspects of traffic. Most of the parameters of traffic flow can be derived from readings of counters activated by the conceptual equivalents of tripwires stretched across the highway lanes (current technology favors electromagnetic induction loops). Furthermore, ground-based instrumentation, once installed, can operate virtually independent of weather and lighting conditions, and in some cases can be used for real-time control. In view of these advantages of ground instrumentation, it appears that aerial photography for observing gross traffic flow should be limited to areas where this would need to be done so infrequently that installing equipment on the ground would not be economical.

Whatever the merits of aerial observation in any given situation, it is an established fact that aerial photographs have been utilized quite widely for observing and studying traffic. This is so in spite of obvious and inherent problems in data collection, including the cost of flying and photographing and the limitations imposed by atmospheric affects, weather and lighting, and in spite of the very considerable effort required to reduce the data to useable form from the photography.

1.2 OBSERVATION OF GROSS ASPECTS OF TRAFFIC FLOW

There have been a number of diverse applications of aerial photography in surveying traffic. One type of application involves

essentially only the counting of vehicles on a stretch of road at a given time.

The work performed by the Freeway Operations Section of the California Division of Highway in the course of monitoring freeway traffic flow in Los Angeles is representative.

The surveillance technique consists of overflying the section of freeway under consideration in a light airplane and photographing successive half-mile segments of the road with a standard 35mm camera equipped with a large capacity film magazine. The data reduction process is rather crude - the photographs are projected on a screen and vehicles on fixed road segments are counted. The road segments considered are typically 800 feet long and are delineated by such landmarks as ramps, overpasses, etc. A surveillance flight may last some 2-1/2 hours, during which a 6-mile section under study would be overflown some 25 times in each direction. The final output for consideration by the traffic engineer is a density chart showing the average number of vehicles per mile per lane as a function of time. From these, travel times can be deduced by utilizing empirically derived relationships between traffic density and average speed.

In this case the requirements of film reading precision are minimal, the only errors of significance being omissions and false introductions of vehicles. For all vehicles except those at the ends of the highway segments the assignment to the correct segment is obvious, and assignment in questionable cases is essentially arbitrary anyway. An inevitable degree of uncertainty regarding car counts in some highway segments arises from the presence of overpasses that obscure parts of the road. In these cases the assumption is made that the vehicle density under the overpass is the same as the average density in its vicinity, and on the average this would be true. Reading errors are estimated by the personnel of the California Division of Highways to vary between 2 -10% omissions; the error rate varies with lighting conditions, increasing sharply as lighting becomes bad.

With current procedures a typical 2-1/2 hour surveillance flight results in some 300 frames of photography, covering some 1100-1200 segments in which vehicles are to be counted. Film reading consists of counting vehicles within each segment and compiling a table of these values. That task currently takes a full man-week of effort. Density calculations and drafting to produce a finished chart take approximately another full man-week.

Currently, no data are being extracted regarding traffic composition by types (cars, trucks). Traffic density by lanes is recorded only in areas where this appears critical. Headway distributions are not recorded, and information concerning vehicle speed is not directly derivable from the photographs, since overlapping photographs closely spaced in time are not taken. These data are of varying degrees of interest, but are not now gathered. The speed data are of particular interest, but these can not be obtained without changing the entire mode of operation.

The improvements in data reduction techniques for such macroscopic analysis of traffic that appear to be most desired are reductions in cost and in time. The additional data that are mentioned above as being of interest are not now gathered because of the clerical burden that obtaining them would impose on the human film readers. Were the film to be read by a computer-controlled system, these data would come essentially at no cost, since a computer, unlike the human brain, is an excellent device for performing bookkeeping functions.

The details of the data gathering and data reduction process involved in observing the macroscopic behavior of traffic by using aerial photography are drived from the work of the Freeway Operations Section, District 7, California Division of Highways. Very similar work has also been done in the San Francisco area, in St. Louis, Missouri, by the Missouri Highway Department, and apparently also in Houston, Texas. It was the strongly expressed opinion of the California highway engineers that continuous congestion inventory of the type that they were engaged in compiling would become recognized as a necessary program of essentially every

highway department concerned with freeway system operation. It was their feeling that the availability of a central service bureau capable of returning fully reduced operations data within a turn-around time of approximately one week would be a welcome service for potential users.

It is, however, a conculusion of the present study that it would not be practical to implement a system to recognize automatically and count vehicles photographed along a given stretch of highway. Programs could readily be written to accept the numbers of vehicles on each highway segment and to produce on a computer-controlled plotter the plots of vehicle density as a function of time and distance along the road. Such a program could be readily implemented at any computer facility equipped with a plotter. It would cut the human effort involved in data reduction and the total elapsed processing time essentially in half.

We consider it impractical to attempt to develop a system using some kind of optical film reader and pattern recognition techniques to recognize and count vehicles on the road segment photographed. The nature of the photography, with each frame differing from the previous one as to the area covered, the orientation of the vehicles within picture, etc., would require very general, and therefore very complex, pattern recognition techniques.

Since a large amount of human intervention would be necessary anyway for each frame to define the highway segments to be considered, and since the pattern recongition task required of the machine would be quite large, such a system could not necessarily be expected to result in major savings in processing time. Its cost would be far greater than present costs. Thus the concept can be discarded without even attempting to estimate the accuracy which such a system might achieve, were it developed. It is reasonable to predict that even at best the accuracy of an automaated system applied to this task would be significantly worse than that of a person simply counting observed vehicles.

1.3 OBSERVATION OF VEHICLE MOVEMENT IN DETAIL

The data reduction tasks discussed above are of the class requiring essentially only the counting of vehicles, with minimal regard for their location. The other extreme in the use of aerial photographs in observing highway traffic occurs when the goal is to measure location very precisely. Work at Ohio State University has involved study of traffic features of the most microscopic kind – the formation and dissipation of platoons of vehicles within a traffic stream. Photographs would be taken while flying above the platoon and staying with it. Some 2500 ft of road would be photographed on a 70 mm frame. Greater than usual care has been taken in reading the film (e.g. – the film was placed between glass plates to prevent warping) and greater accuracies have been achieved. Errors in position of as little as 6 in and velocity errors of less than 1 m.p.h. have been claimed. The results have been achieved at a considerable cost in time. The films have been read with an analytical stereoplotter – an accurate but slow device, since the average reading rate seems to have been 20 sec/point. Such performance appears beyond the resolution power of any but the ultraprecise and very slow automatic scanners. No practical purpose could be achieved by using such devices for this task, since the time to examine a frame of photography would in fact increase were one of these mechanical scanners employed.

In our judgement the techniques of automatic pattern recognition can be both practically and profitably applied to those film reading tasks that are highly repetitive, do not require much human intervention and involve the measurement of many vehicle locations without burdening the automatic equipment with a difficult pattern recognition task in each case. The projects using data from aerial photographs of traffic that will be described below are considered as representative of the class for which automated data reduction could be practical. There have been a number of projects of this class carried on at various institutions. Only those details of the research that affect the accuracy requirements of the scanning process are relevant to a discussion of scanning techniques.

The System Development Corporation conducted a study of traffic flow through a diamond interchange. The interchange was photographed from a helicopter hovering at an altitude of some 2100 ft. A Maurer 70 mm camera was used with a Zeiss Biogon 72° lens. This gives ground coverage of about 4000 ft on the diagonal, of which about 3000 ft were considered in the study. To assure full coverage, it was considered necessary to photograph an area extending some 500 ft on each side beyond the area of interest, since a helicopter may experience difficulty in maintaining station and attitude. Photographs were taken at 1-second intervals. Faster repetition rates were not considered practical, largely because of expected mechanical problems of camera wear, etc.

The model of a diamond interchange used by SDC involved dividing the interchange into functional blocks, such as ramps, approaches to ramps, segments between off and on ramps, etc. The ground reference points of interest were, therefore, the boundries of these functional blocks. They were marked on the ground by 8 ft x 1 ft strips placed on the shoulders of the roadway.

Matters of interest to the researchers were the time each vehicle spent traversing each block, its speed in crossing block boundaries, and queue lengths as a function of traffic signal states. The traffic signal state at any time could be seen in the photographs from special vertically pointed lights, installed for the purposes of the experiment.

The Institute of Transportation and Traffic Engineering at UCLA conducted a study of highway exit ramps on lane changing behavior. The phenomena of interest were lane changes and their context - i.e., distance from the ramp, traffic speed, and traffic density, particularly gap length in the lane being entered. Seventy mm photography covering 1 mile of road was used. The researchers felt that the 1-mile section considered was probably too short - they would have prefered as much as 5 miles. Such data are out of reach, however, as much because of the inability of a helicopter to hover at sufficient altitude as because of film reading problems that might be encountered trying to cope with photography of that scale.

In both these studies there are several features that make it appear reasonable to automate the data reduction task. The site photographed was essentially the same from frame to frame (varying only due to the inability of the helicopter to be perfectly stable and stationary), so that the area of interest would not need to be delineated afresh for each frame. The desired output is quite extensive for each frame of photography, in any case practically requiring computer manipulation. Going from one frame to the next, it is known that most vehicles seen in one frame will again be visible in the next within some reasonable interval from their previous position; thus in most cases there is prior knowledge which can be used to greatly simplify the search and recognition process. Finally, the sheer volume of work is such that any reduction in the time consumed and the work expended is likely to be significant.

The current data reduction technique is to have the flim read by operators operating a Benson-Lehner Telereadex device. This is essentially a table on which the film is projected at either 10X or 20X magnification. (The operators use 20X for this task.) The table is equipped with a pair of crossed wires, one horizontal and one vertical, which the operator can move by twisting a handle with each hand. When she depresses a foot switch, a digital encoding of the positions of the wires at that instant is recorded. At the same time a sequence number, incremented automatically for each vehicle, a film frame number, incremented automatically when the film is advanced, and various other codes, set by hand on a number of dials, are also recorded. The reader is equipped with an elaborate film advance mechanism allowing foreward and backward advancing of the film as well as rotation of the projected image - convenient in that it is generally used to align the roadway with one of the cross-wires, so that only one wire need be moved when measuring the locations of successive cars in a line.

2. AUTOMATED FILM READING

2.1. PRACTICAL CONSTRAINTS

An experienced operator can read car locations quite rapidly. Even so, in the case of the SDC study, the quoted rate for reading film was roughly 20 hours for the 60 frames (each with some 500 cars) representing one minute of traffic. The position recording rate quoted by UCLA was 15 - 20 minutes per frame (each with some 150 - 200 vehicles). This appears to be slower than the SDC rate, but the difference is probably explainable in terms of the smaller scale of the photographs. It should be noted that in these systems no provision was made for recording the vehicle type (car, bus, truck), and no descriptive data (such as color) were recorded which might subsequently have been an aid in linking up vehicle trajectories. Including such data would have imposed a burden on the operator further slowing the work. Even at the rate quoted, reducing the data corresponding to one hour of traffic photographed every second requires about half a man-year of effort. The desirability of automation to speed up the process is obvious.

As an ideal solution, one might seek a system which would perform the following functions automatically:

1. Perform image enhancement and other transformations required to suppress background and noise.
2. Impose a suitably rectified reference grid on each picture frame.
3. Delineate the boundaries of highways, bridges, interchanges, intersections, and other areas of interest.
4. Locate each vehicle, record its coordinates, and make measurements to be used for identification.
5. Apply automatic pattern recognition techniques to these measurements to classify each vehicle as a car, bus, etc.
6. From the data on two successive frames, compute the displacement and instantaneous velocity of each vehicle.

7. From data on several successive frames, compute and verify vehicle trajectories, and calculate such traffic characteristics as vehicle separation, passing frequency, distribution of vehicle lane changes, and location of bottlenecks.

8. Generate suitable displays and printouts.

When specifying a system to be actually implemented, practiality demands some deviation from this ideal. Clearly, an automatic system that accepts pictures of an unidentifed area void of annotation, identifies landmarks in the picture, and then scans subareas of interest to the user can not be devised.

Several techniques could be employed to transfer such positional information to the computer. The simplest, in the sense of requiring the least equipment, would be to place appropriate markings directly on the film. It appears to be relatively easy for a computer to locate large opaque markings on the film. This approach has drawbacks. Marking the film is a somewhat tedious process and difficult to do with both precision and speed. Furthermore, the approach implies hands-off operation once the film is in the scanner. This is undesirable, since any error by the human in marking the film or any misinterpretation of the markings by the computer would go undetected, or, if detected, could not be conveniently corrected.

2.2 INTERACTIVE FILM READING SYSTEM

Our system is designed to allow imput of road delineations and landmarks directly to the computer in a manner allowing immediate verification. It is implemented by a display on-line with the computer and a graphics tablet. We display the picture scanned at rather coarse resolution on the CRT display and superimpose on the display a marker corresponding to the location of the stylus on the graphics tablet. Routines are provided for selecting small subareas of the photograph being scanned for display in greatly enlarged format at high resolution and for selecting reference points within these high resolution sections.

The following steps are envisioned in the operation of the completed system in scanning a set of successive photographs of substantially the same area:

1. The first frame is displayed in coarse resolution. The operator causes the neighborhood of each ground reference point to be displayed enlarged and at full resolution, and marks the landmark precisely. The ground coordinates of the landmark are supplied to the computer via teletype or other convenient means.

2. The operator then traces the shoulders of any roads in the picture to be scanned. The computer scans in the immediate area of the indicated shoulder, locates the discontinuities in the film presumed to be the shoulder and computes and stores their ground coordinates.

3. All vehicles in the first frame are located interactively. The operator indicates by graphics tablet the position of all vehicles, and these are recorded by the computer. Using contour tracing algorithms, the computer finds the vehicle images at the indicated positions. It then makes measurements to classify each vehicle as a car, bus, or truck. A copy of each vehicle image is stored for future use.

4. The computer estimates the position of each vehicle in the next frame. Also, areas of the picture are located where vehicles not present in this frame may appear in the next frame.

5. The next frame is scanned.

6. The fixed landmarks are located either by the operator as before or automatically by the computer. In the latter mode, the computer starts searching for each landmark at the position it appeared in the previous frame. Template matching is used to find a match. With the landmarks known, the coordinate transformation between the previous frame and the present frame is computed. Then the road-edges can be located automatically by performing the transformation from the road-edges in the previous frame.

7. Vehicles that appeared in the previous frame are located automatically as follows: The computer searches for each vehicle starting at the predicted position. Possible matching vehicles are found and template-matched against the stored vehicle. The operator is called only if the computer fails to find a match. When found, the vehicles are stored as in Step 3.

8. The operator searches in the picture areas where vehicles that are not in the previous frame may appear and indicates all such vehicles. The computer treats these as in Step 3.

Steps 4 through 8 are repeated until all the frames have been processed. The successive locations determined for each vehicle are stored. Finally, the data of interest are abstracted and output in a form directly useful to the investigator. It is probable that in some cases routines can be developed that achieve savings in the amount of necessary scanning by determining that some data are of no interest in the given application. For instance, one can envision that for studies evaluating the effect of trucks on traffic behavior, the computer would select for scanning only those sections of road in the vicinity of trucks, etc.

The central idea leading to this choice of an approach was that pure pattern recognition - the recognition that some small area of the film being scanned contained the image of some vehicle - was too difficult a task to impose on a computer to be practical. A number of simplifications of the general task accrue from our approach. First, the fact that initially each vehicle is identified as such by a human operator eliminates the problem of identifying an arbitrary shape at an arbitrary position as a vehicle. A human operator in making such an identification implicitly performs a very complicated decision making process, guided by experience and reasoning power. It may be instructive to enumerate some of the decisions that a human operator performs implicitly and that an automatic device would have to be programmed beforehand to perform explicitly:

a. Is there an object in the area being considered that is darker than the background?

b. Is there an object lighter than the background?

c. Is the object casting a shadow ahead, behind, or to either side?

d. Is the object a small car?

e. Is the object a large car?

f. Is the object a truck or bus?

g. Is what appears to be an object not a vehicle after all (but instead, e.g., an asphalt patch, a shadow, a spot on the film)? Consider the complicating factors:

a. The vehicle-sideways-shadow images of vehicles in adjacent lanes may merge.

b. Vehicles may be quite close together within a lane (tail-gating), so that again several vehicle-shadow images merge.

c. Vehicles may be non-uniform in shadow (Truck cabs and bodies, two-tone cars, light cars with dark vinyl tops, etc.).

d. Vehicle images may be intersected by shadows falling across the roadway.

e. Vehicles may appear in different orientations with respect to the picture coordinates and even with respect to the edge (while changing lanes).

f. While reasonably constant from one frame to the next, such parameters as picture scale (depending on camera altitude), film contrast and exposure level, and length and direction of shadows can be expected to vary from instance to instance when observations are made.

These lists of complications should make clear that any attempt to recognize visual patterns as vehicles on the basis of

comparison against some set of archetypical vehicle images is likely to encounter great difficulties. A substantial number of archetypes would have to be defined to cover the range of possibilities. Even then, any given vehicle might deviate sufficiently from even the closest archetype to present a difficult problem of decision-making.

Generally speaking, enlarging the set of standard or archetypical patterns against which a given unknown pattern is to be matched should allow better matches to be achieved on the average, since the typical patterns can be chosen to represent more nearly similar objects. On the other hand, the average number of comparisons that have to be performed before a given object is either satisfactorily identified as a vehicle or rejected would grow with the number of archetypes. Furthermore, the selection of the standard patterns for comparison is by no means a trival task in itself. These considerations apply whether the comparison is to be performed by a computer on the images recorded digitally or on some abstracted measurements or features of the observed pattern, or whether it is to be performed optically by coherent light filtering techniques.

The technique proposed here has the merit that only one "standard" pattern need be tested in searching for any given vehicle, a standard based on the particular vehicle itself with any and all pecularities that it might have, degraded only by the noise inherent in scanning it.

2.3. ALTERNATIVE SYSTEMS

Conceptually attractive as an alternative is the method of observing differences between successive film frames. Any contigous area of appropriate size in which a sufficient change in overall film density was observed is assumed to be an area in which a vehicle was located at the time of one of the frames was exposed, and not at the time the other was exposed. Such a system has not been tried out experimentally, but a number of observations about the concept can be made.

If the camera is held absolutely fixed from frame to frame, then any given point on the ground will always register on the same point of the film and will be read by any film reading device at the same pair of coordinates, subject only to misalignments due to irregularities in advancing the film in the camera and the film reader, and in variations and distortions of the film itself. These misregistrations should be relatively minor; in a system in which the difference point by point was to be computed digitally in a computer, appropriate offsets could be computed to bring fiducial marks into alignment. Misregistration might be one problem, among others, in a system which relied on analog circuitry for determining differences. (Such a system might attempt to subtract the signals obtained by video scanning a frame of film and by reading some video recording of the previous frame, recorded either on a tape or a storage tube).

In the case under consideration here, the camera is not held stationary but is subject to lateral and vertical translation, rotation, and tilt as the helicopter moves. To achieve cancellation of those features in the images which have remained invariant on the ground, appropriate transformations have to be performed on the picture coordinates to assure that the same ground points are brought into coincidence. To observe the magnitude of these effects, consider a helicopter hovering at 3000 feet and photographing an area one mile square. Assume that the initial photograph is taken at the nominal position and vertically. Consider now the following changes in helicopter position and altitude in the interval before the next picture is taken, and their attendant effects on picture registration. These will be expressed in terms of apparent displacement of a point horizontally on the ground - i.e., the distance which an object on the ground would have had to move to appear on the film at its position in the second frame, had the camera been stationary.

a. The helicopter changes altitude by 15 inches. Points at center of the picture are unaffected. Points near the edges of the picture are apparently displaced one foot radially.

b. The helicopter moves laterally. All points are apparently displaced the same distance. A displacement of several feet per second is to be expected.

c. The camera is rotated about the lens axis due to helicopter rotation. The effects increase going radially outward from the center of the image. A one degree rotation will result in a virtual displacement of 17.5 feet for a point 2000 feet from the center of the picture.

d. The camera is tilted 1 degree. The point immediately below the helicopter will undergo virtual movement of 6 feet. A point 2000 feet from the center in the plane of the tilt will be virtually displaced 76 feet in the plane of the tilt.

It is clear from the above that two successive frames can not be compared by any technique conceptually equivalent to directly overlapping one over the other. Instead, it is necessary to derive precisely the mapping function between the film coordinates for each frame (a process known as resection) and then to apply it point by point over the area of interest. It may be noted that relatively small changes in camera orientation bring about large changes in the mapping of points from the ground to the film. It follows that significant errors in the mapping function may be introduced by relatively small errors in estimating the camera position and orientation.

Image resection is based on a number of observed ground points in excess of the theoretic minimum of three required to fully determine the ground-to-film mapping function. The excess points, of which there may be some ten, tend to reduce the effects of random errors. Even so, it appears somewhat ambiguously from reports on work performed at UCLA that random errors of the order of 1-3 feet were observed in testing the mapping of known points on the ground. Hence, if in succeeding frames the errors happen to be in opposite directions, significant mis-alignments can occur, giving rise to spurious difference values.

Even if the mapping is perfect, the computed difference between supposedly identical picture pairs will not normally be zero because of noise effects in measuring the film density in the two frames. Assuming that each measurement at a picture point is subject to additive random uncorrelated noise, with a normal distribution and a standard deviation σ, then the difference values will be subject to such noise with a standard deviation $\sqrt{2}\,\sigma$.

Assuming that the pairwise differences between successive frames of imagery were computed, a given moving vehicle would be registered (as an area of detected difference) at a given location in two successive frames, and at two different locations within each given frame. In particular, it would register in its current location as a difference between its actual image and the image of the presumably empty road at the previous instant of observation, and at its previous location as the difference between the road segment that it vacated and its image at the previous instant of observation.

Any trajectory matching algorithm would have to be capable of sorting out the two kinds of vehicle appearances and also for accounting for various observations, such as partial or complete overlaps between two kinds of difference images. In particular, it should be noted that a stationary vehicle will not be visible in the differenced pictures, and vehicles closely following each other might effectively obliterate each other. (These last problems would make this approach difficult to apply to the diamond interchange study mentioned above). No current vehicle trajectory matching programs exist that are intended to sort out such differenced images. Because of this consideration, and because of the other problems enumerated - the substantial calculation involved in bringing such pictures into proper alignment, and the anticipated difficulty in detecting and delimiting the "areas of real difference" in the noisy array of difference values – this approach to vehicle following was not investigated in any further detail at TSC.

Similarly, the idea of using coherent light optical correlation techniques to detect and identify vehicles was rejected without experimental investigation for a number of reasons. This approach rests on the following physical and mathematical bases: A spherical lens performs a two-dimensional Fourier transformation, in the sense that if there is variation in intensity and phase across a collimated beam of coherent light falling on the lens, the lens will produce in its focal plane a distribution of light that is the two-dimensional Fourier transform of the illuminating light distribution. A second lens will produce the inverse transform. These observations can be utilized to perform filtering or correlation as follows: The image (transparency) to be processed is placed in a collimated beam of coherent light from a laser to create intensity modulation across the beam. A frequency domain filter can then be placed in the focal plane of the lens to perform filtering. The filter will attenuate various frequencies in proportion to its density in the corresponding area. For instance, a high-pass filter would be opaque in the center (low frequencies) and clear elsewhere; a low-pass filter might consist of a small clear hole on the optical axis and be opaque elsewhere, etc. It is also possible to perform phase filtering, by varying the optical thickness of the filter. The second lens, which performs the inverse Fourier transform and in the absence of any filter merely restores the original image, will, with the filter present, see a modified Fourier transform and will therefore produce an image with its spatial frequency content selectively modified.

In the application at hand, the spatial filter to be used would be designed to produce the cross-correlation of the given aerial photograph with a vehicle image. In general, convolution in the space domain is equivalent to multiplication in the spatial frequency domain. Thus, if F [ ] denotes Fourier transformation, and $i(x,y)$ is an image and $f(x,y)$ the impulse response of a spatial filter, $F[i(x,y)] = I(\omega_x,\omega_y)$ and $F[f(x,y)] = F(\omega_x,\omega_y)$.

If * designates convolution

$$i(x,y)*f(x,y) = \int_{\infty}^{\infty}\int_{\infty}^{\infty} i(x-\xi,y-\gamma)\ f(\xi,\gamma)\ d\xi d\gamma$$

then

$$F[i(x,y)*f(x,y)] = I(\omega_x,\omega_y)F(\omega_x,\omega_y)$$

If the frequency domain filter $F(\omega_x,\omega_y)$ is selected to be the transform of a vehicle image, then the cross-correlation between the given aerial photograph and this vehicle image will appear in the output plane of the system. Hopefully, the vehicle image will correlate well with the images of all vehicles of that type visible in the photograph and poorly with everything else, so that the output will be a number of bright dots at the position of the vehicles against a uniformly dark background. Furthermore, this correlation function is virtually available instantaneously. It is this latter point that is most appealing about this approach and is most stressed by its proponents.

Unfortunately, certain practical difficulties must be taken into account in considering the implementation this approach in a system. In the first place, the locations of the correlation peaks would still have to be measured in the output plane, and the corresponding ground coordinates would have to be determined. Thus the output plane would have to be availabe for high-precision scanning both with the filter in place (to create the correlation peaks) and without the filter (to enable the locations of ground reference points to be measured, since these would not be directly observable in the filtered view). In addition, because of the variability of possible vehicle images (as discussed above),a number of different filters would have to be applied in succession if reasonable correlation values were to be obtained. This implies both repeated scanning (counteracting the merits of "instant" correlation) and repeated replacement of filter. The effective physical size of the focal plane filters that would be used is a function of the focal length of the transforming lens and of the wave-length of the laser; in a practical case, these would amount to a few milli-

meters. Thus the need for precise mechanical alignment and the problems this is likely to cause are evident.

In summary, the approach of usint coherent light correlation techniques was rejected because of the obvious increase in hardware system complexity it would entail, and because in the final analysis its success would depend on highly reliable matching of prespecified sample patterns against a widely variable population of observed patterns amidst noise. The chances of success were not judged good.

3. HARDWARE IMPLEMENTATION

The experimental work performed in the course of this program utilized the TSC TRansportation IMagery (TRIM) system shown in Figure 1. The overall system consists of a computer controlling a flying-spot scanner and a CRT display, and receiving inputs from the flying-spot scanner and a graphics tablet.

3.1 FLYING-SPOT SCANNER

The key item in the system is the flying-spot scanner. At the present state of technology, a flying-spot scanner appears to be the most suitable film reading device available for the film reading necessary for tasks of the type considered here.

In concept, computer-controlled flying-spot scanner is a fairly simple device. Its principal parts are, in addition to the computer, a cathode ray tube (CRT) with its associated electronics, optics, a film holder and a light measuring device, normally a photo-multiplier tube (PMT). The system operates as follows:

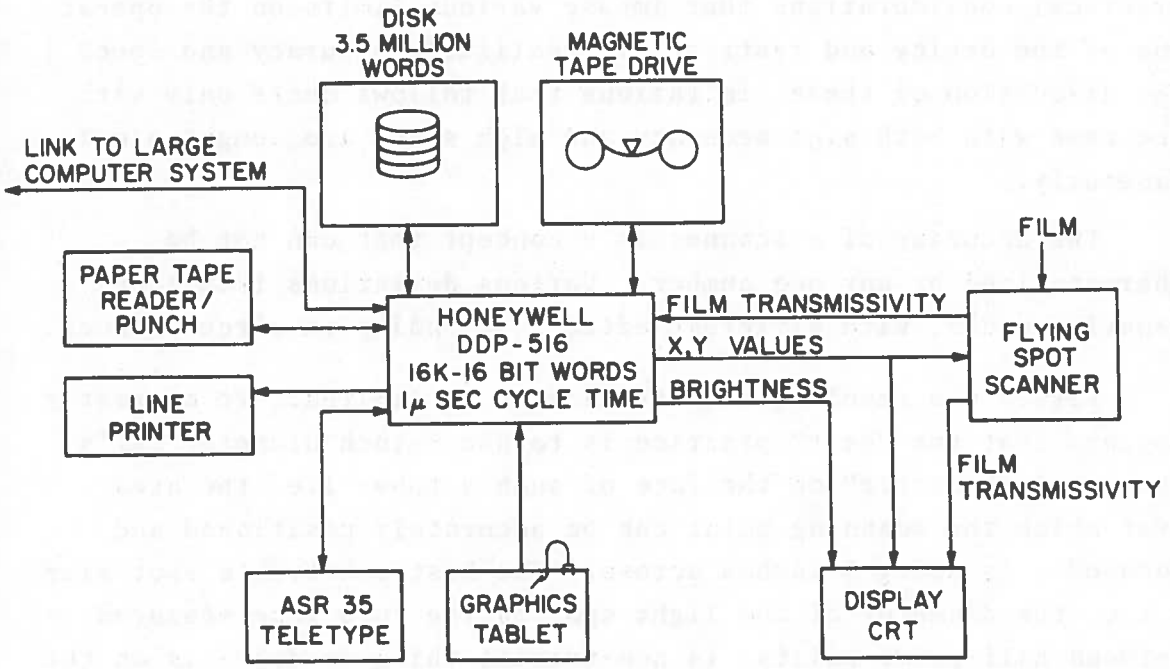


Figure 1. Transportation Imagery System

The computer outputs to the scanner the x and y coordinates of a point on the film to be read. The digital signals of the computer are converted by D/A converters and suitable amplifiers to electrical signals which deflect the electron beam of the cathode ray tube to the appropriate point on the tube face, creating there a point of light. This light is gathered by a lens and focused onto a corresponding point on the film. Depending on the density of the film at this point, a certain fraction is transmitted through the film. This is collected by another lens and directed at the photomultiplier tube, which produces a proportional output current. This current is measured; the measurement is converted to digital form by an A/D converter and read by the computer as input. The computer can then output the next set of x, y coordinates to read the film density at some other point. In particular, the computer may use the density values obtained at the set of points previously read to calculate the next point to be examined. This is an important capability, since it may in certain cases greatly reduce the total number of points that need be read.

The outline of operation given above omitted a number of practical considerations that impose various limits on the operation of the device and restrict its realizable accuracy and speed. The discussion of these limitations that follows deals only with the case when both high accuracy and high speed are sought simultaneously.

The accuracy of a scanner is a concept that can not be characterized by any one number. Various deviations from ideal behavior occur, with different effects depending on circumstances.

First, the resolution of the device is limited. It currently appears that the "best" practice is to use 5-inch diameter CRT's. The "quality circle" on the face of such a tube- i.e. the area over which the scanning point can be accurately positioned and focused - is about 3 inches across. The best achievable spot size - i.e. the diameter of the light spot on the tube face measured between half-power points (a non-trivial thing to do) - is on the order of .0005" - .001". This means that a grid of 4000 x 4000

reasonably independent points is about the maximum that a scanner of this type can read. Scanners can be built with more addressable points on the grid (for instance, Information International, Inc. has built scanners with 16,000 x 16,000 addressable points), but the adjacent addressable points largely overlap. Additional degradation in resolution comes about because of the optics and even because of scattering in the film being read itself. Resolution of optical systems is conventionally expressed in terms of the modulation transfer function (MTF) of the system. The MTF is defined as the gain of the system as a function of spatial frequency (cycles/mm), normalized to one at frequency zero. The apparently best performance curve achieved shows an MTS of 0.95 at 30 cycles/mm, 0.50 at 50 cycles/mm, and 0.15 at 100 cycles/mm. These performance figures represent the upper limit on achievable resolution. It should be appreciated that such performance is achieved by employing considerable sophistication. Dynamic focusing is used to focus the electron beam in the CRT (i.e., the magnetic field for focusing the beam is adjusted to compensate for changes in the length of the electron path when the beam is moved off axis). The objective lenses are specially designed to be color corrected for the particular phosphor used in the tube and to compensate for the light bending efficts of the glass face-plate of the CRT (which is about 1/2-inch thick, for mechanical stability).

The positional accuracy of the spot is also limited. Despite "geometry correction" circuitry, a mathematically perfect rectangular grid of points is not achievable. Some "pincushion" or "barrel" distortion remains, and in addition, the x and y axes may not be perfectly perpendicular to each other. These systematic effects can be kept well under 1%, however, and to the extent that they are still significant, can be lumped with the various other distortions that appear in the mapping from ground coordinates to perceived film coordinates and removed by the same set of calculations.

Somewhat more difficult to cope with is the problem of hysteresis. The demands of precision in beam deflection and focusing are such that they can not be met with electrostatic deflection

plates and focus devices, but rather the deflection and focusing must be done by applied magnetic fields. Even though the magentic core materials for the coils are chosen to have low residual magnetism and narrow hysteresis loops, hysteresis effects can not be avoided completely. Thus, the precise position to which a spot is deflected will depend upon the previous scanning pattern as well as upon its nominal coordinates.

The maximum possible positioning error due to hysteresis apparently can not be kept under 1 part in 1000 by design of the hardware (i.e. about 4 spot diameters). If care is taken in programing, however, its effects can probably be effectively eliminated. To this end, any given area of the film must always be approached from the same general direction.

Second-order effects limiting performance seem to abound with any precision equipment. With magnetically deflected scanners, the effects of heating also fall in this class. Since the deflection coils carry substantial currents (several amperes) at large deflections off axis, they can undergo mechanical deformations due to changes in temperature. These change the effective magnetic fields and therefore the location of the scanning spot. No quantitative estimates for this effect are available. Careful design probably can reduce it. In any case, it is true of both this and the hysteresis effect that while they may affect the absolute accuracy of the measurement and its repeatability, they would not ordinarily affect significantly the measurements of relative placement of objects close together in the film.

In addition to a certain randomness in the positioning of the scanning spot, there is also some randomness inherent in the measurement of the film transmissivity at that spot. The transmissivity of film is the fraction of incident light that passes through it. Since there is considerable variability in the amount of light emitted by the CRT phosphor as a result of a perfectly constant electron beam bombardment ($\pm$ 10%, randomly distributed over the tube face, and changing as the tube ages), it is necessary in precision scanners to measure both the incident and the transmitted light and compare them to get the true transmissivity. To

this end, a beam splitter is inserted in the light path between the CRT and the film. A fixed fraction (~ 10%) of the light emanating from the CRT is measured by a reference photo-multiplier tube (PMT). The rest is directed through the film, and the transmitted part is measured by the main measuring PMT. The quotient of the two readings is the transmissivity of the film. The logarithm of this is the quantity known as density. Frequently the measurement made is that of density because the human eye seems to preceive logarithmically equispaced steps in light as "equal steps of brightness," and because a better dynamic range can be obtained.

The values read as the intensity of light at either PMT are random variables in the true statistical sense of the term. At the low light levels obtained with flying-spot scanners, the fact that light is a stream of discrete photons arriving at random intervals assumes practical signficance. One can no longer think in terms of the instantaneous value of light, but must think instead of an average value over some interval. From statistical theory it follows that the variance between the observed average and the "true average" value will decrease as the period over which the average is computed is increased.

The precise interval needed to achieve a certain mean square error varies with among other things the beam power of the CRT, the phosphor efficiency, the aperture and focal length of the objective lens, and the efficiency of the PMT. In practice it appears that an integration interval on the order of 5 μsec is necessary to get reasonably reliable gray scale resolution to about 32 levels. This is approximately the gray scale resolution capability of the human eye at normal ambient light levels.

The speed capability of the flying spot scanner can be defined in terms of the length of the interval between the instants at which the gray scale levels of successive points can be read by the computer. During this interval the computer must perform certain housekeeping chores: store the last light value read, obtain new value for the coordinates, output them, and read the new value of the gray level. The scanner must deflect the beam to the specified location, allow an appropriate interval for the integration of the

light and provide for the analog to digital conversion. Some of the functions of the computer and the scanner can overlap in time. For reading film in a raster scan mode with x incremented by fixed amounts and y constant, a DDP-516 computer (a machine falling in the upper part of the class referred to as mini-computer) requires a total of 14 memory access cycles of 1 μsec each per point. Within this cycle the interval between the output of the latest x-coordinate and the reading of the gray level is 9 μsec. This is adequate time for the scanner to perform its functions, allowing 2 μsec for A/D conversion.

Thus, the data transfer rate to the computer in this mode is roughly 7×10^5 values/sec. This compares favorably with the highest speed of data input to the DDP-516 via magnetic tape. The highest speed tape drive available for it operates at 80 i.p.s. with 800 b.p.i. tape, giving a character transfer rate of 6.4×10^5 characters (6 bits)/sec – about 10% slower than the scanner.

The specific number cited above pertain specifically to the TRIM system, but little deviation is to be expected for other systems operated in a similar manner. It may be noted that the computer is fully occupied during the entire scanning time in performing "housekeeping chores" –computing the coordinates for the scanning spot and the location for storing the value to read, checking for line ends, etc. Relatively simple additional hardware could be incorporated in the system to free the computer of this burden and allow it to perform other operations while a picture segment was being scanned, while at the same time increasing the scanning speed. Such a scan controller would consist of digital circuitry, largely computer interface logic and two counters, and would perform the following operations: It would accept from the computer the horizontal and vertical limits of the film segment to be scanned and then automatically generate with its counters the deflection signals for a raster scan of that area at an appropriate rate. The film density measurements would be performed automatically in the scanning sequence, and the digitized value would be entered into computer memory via a direct memory access port without interfering with the operation of whatever program steps the computer might be executing.

Flying-spot scanners can read color film. Scanners designed to do this must use a CRT with a "white" phospor (generally P24) and some system of colored filters. Generally, the design is like that of the TRIM scanner at TSC – a wheel with colored filters mounted in it is placed in the light path between the CRT and the film. In operation, a picture is read in turn with each of the three color filters in place. The filters must be moved mechanically whenever a different color is to be read. This arrangement has several advantages: First, only 2 PMT's (reference and reading) with their associated electronics are needed. More important, the device can also be used for output in color. If unexposed color film is inserted in place of the transparency, the color filters between the film and the CRT allow the film to be exposed in turn to the three different colors to produce a color picture. On the other hand, scanners of this configuration become impractically slow because of the required mechanical movement of the filters if one should want to obtain a full-color representation of an individual small area before proceeding to scan the next small area, again in full color. This, unfortunately, is precisely the scanning mode useful for scanning traffic imagery.

The TRIM system flying-spot scanner has a color capability of the type described. Only limited trials were made using it, largely because of the penalty that such use would involve in terms of speed of operation. All scanning and any operations on the scanned pictures would have to be performed repeatedly, storage requirements would increase three-fold, etc. Furthermore, since the color filters inherently reduce the effective light level in the system, the readings with filters in place resulted in lower and noisier readings. Vehicle contrast against the rather neutral gray background of the road surface was not enhanced by the colored filters.

Although it is intrinsic to the use of color-separated digitized images that more processing will be necessary than if only single (one color or unfiltered) images are used, it is possible to construct a flying-spot scanner in a configuration that does not necessitate mechanical movement of filters and repeated scanning

to obtain the color separated representations of images. A CRT with white phospor is used, and filters are placed between the CRT and the film (thus color output is ruled out). The separation of light into component colors is performed upon the transmitted light by a set of dichroic mirrors.

A dichroic mirror is essentially a beam splitter which reflects light in a certain spectral range and transmits the rest. Two different mirrors can be used to deflect the transmitted light in two spectral ranges to off-axis PMT's. A third PMT can be placed on axis to measure the remaining light. The operation of such a system then involves the reading by the computer of three different values representing the three different colors after each positioning of a scanning spot. Assuming the same equipment speeds and scanning mode as used before in the calculation of the reading cycle time for the TRIM system, we find that to input and store 3-color data would require 22 μsec per point, as opposed to 14 μsec for only one transmissivity value per point. With a small amount of additional circuitry, the scan controller described above could also drive such a system with no increase in elapsed time for scanning an image in three colors over scanning it in one.

3.2 OTHER FILM SCANNING DEVICES

Beside flying-spot scanners, a number of other devices can be used to scan film. For completeness of this section, they are briefly discussed below. They are not judged suitable for the task under consideration for the reason indicated.

3.2.1 TV Cameras

These employ camera tubes such as vidicons, image orthicons, etc. All these have photosensitive surfaces on which the image to be scanned is projected. Electric charges gradually build up on such a surface as it is continuously illuminated. These charges are drained off point by point and the currents are measured to obtain a measure of illumination at the point. Since the charge read off is a function of both light level and length of time since it was last read, random scanning under computer control is

not possible, but instead an image must be read in a regular raster scan mode at a fixed rate. This is not a desirable mode of operation for the problem at hand.

3.2.2 Image Dissector Cameras

An image dissector tube has a photo-emissive front surface onto which the image to be examined is projected. The emitted electrons are drawn by an electrostatic field backward in essentially parallel paths, such that the current density at any cross-section of the the beam is proportional to the distribution of illumination on the tube face. In concept, the back of the tube is a conductive plate with a small hole in it. Most of the electrons in the beam hit the plate. Those corresponding to one small area on the photo-emissive surface pass through the hole. There they are collected, and the resulting current is measured. The whole beam emanating from the front surface is deflected so that the current from different spots on the tube face is caused to hit the hole in the back. The deflection circuitry is of the same type as that in CRT's, and the needs to integrate the photoemitted currents to get meaningful average values are the same as in the case of flying-spot scanners, so that operating speed can be roughly the same for both devices. The disadvangage of image dissectors versus flying-spot scanners seems to be that currently it is possible to obtain resolution several times better with a flying-spot scanner than with an image dissector, and that with image dissectors there appears to be no way to avoid using mechanically moved color filters to obtain color information about the input image. Image dissectors use light from an external source and can therefore analyze opaque images (rather than transparencies) as well as real world scenes. This would make an image dissector a suitable scanning device in a system using coherent light optical filtering, which a flying-spot scanner would not be.

3.2.3 Mechanical Scanners

Various mechanical scanners are in use. These all use mechanical motion of either the image or a scanning head to position the scanning spot. Some light source is focused on the spot

to be read, and the light is then measured by a photomultiplier. The range of such equipment goes from facsimile news photo scanning equipment to micro-densitometers. The equipment can be made extremely precise in every way and can at least theoretically read images of any size. Again, it does not appear suitable for the purpose under consideration here because it is too slow for random accessing of points under computer control, and line-by-line raster scans are not a desired computer input.

3.2.4 Laser Scanners

Various scanners using a laser as the primary light source have been proposed, and quite probably some have been built. They promise very high accuracy, but as yet it would appear that no satisfactory method of random point accessing is available.

3.3 REQUIREMENTS FOR SCANNER IN AN OPERATIONAL SYSTEM

The flying-spot scanner in the TRIM system was built to read 35 mm film. Its film holder mechanism is primitive, and the automatic film advance under computer control is not fully implemented. It would not meet the following requirements of an operational system: The scanner should be equipped to read 70 mm film; the film advance mechanism should be computer actuated. Spatial resolution should be such as to allow readings to within no worse than one foot on the ground. Since ground coverage of an area at least a mile across is a practical requirement, the scanner would need to resolve at least 5280 points in 70 millimeters, or 75 points/mm. With the very best flying-spot scanners, reasonable resolution can be obtained at such spatial frequencies. As a practical matter, a grid of some 8192 x 8192 addressable points would have to be provided ($8192 = 2^{13}$). This, for photographs covering an area a mile square, would correspond to addressable points about 8.5 inches apart on the ground. Such points could not be clearly resolved individually but would give a clear picture overall.

The gray level resolution of the scanner should be to 5 bits (32 levels) and the scanning speed in the raster scan mode should be the fastest possible consistent with the other requirements. A scan controller generalizing the raster scan pattern and entering the values read into the computer via a direct memory access channel should be provided. The effective reading rate should then be on the order of one point every 10 microseconds, with the computer free for other tasks.

3.4 COMPUTER SYSTEM REQUIREMENTS

The computer in the TRIM system is a Honeywell DDP-516, with a memory of 16,000 16-bit words. It is one of the larger of the small computers. There are a number of other competitive machines of that class which could serve as well. It appears likely in view of the considerations below that a smaller computer could do the job as well if it were given access to a large time-sharing system for some of the required calculation.

With the current system, the computer spends much time in performing scans and in outputting images to a display. The external scan generator would free it of that burden. The other major loads on the computer are the matching of images by cross-correlation, the tying together of vehicle trajectories, the various calculations in performing the mapping from film (scanner) coordinates to ground coordinates, and various "housekeeping" functions, of which the principal one is storing on disk the vehicle images found in a given frame and retrieving from disk the images of vehicles from the previous frame.

These tasks, with the exception of the resection task involved in the coordinate mapping, seem appropriate for a dedicated small computer. They are within the capabilities of a small computer, and they would not constitute a cost effective use of a large time-shared machine.

The cross-correlation of images stored in digital form as matrices of numbers of five significant bits is well within the capability of a 16-bit machine. If performed on a large machine,

the task would tie up much more elaborate (and much more expensive) equipment for a comparable length of time. Because the task is basically simple, but highly repetitive (it would be performed more than any other) the routines involved should be programmed in assembly language for maximum efficiency.

The task of extending vehicle trajectories, first by extrapolating a given trajectory to obtain the next likely location for the vehicle and then by adding the next point once the actual vehicle location has been found, is quite complicated, requiring many logical tests for various conditions and contingencies. It requires a quite elaborate program, but not in itself much actual calculation. As a practical matter, because of its complexity it has to be written in some higher-order language (e.g. FORTRAN) for comprehensibility. The routines should readily fit on a computer of the DDP-516 class.

The disk file management system required to keep track conveniently of the various vehicle images and also the vehicle trajectory records is probably somewhat different in kind from the disk operation programs usual on minicomputers. Its requirements certainly exceed the capabilities of the TOP operating system in use on the TRIM DDP-516. The requirements are not particularly complicated, however, and a suitable file handling program for a minicomputer should not be difficult to write, probably as a combination of FORTRAN and assembly language. The number of disk accesses required during operation can be estimated. For each vehicle in a frame, the disk record containing its trajectory would need to be read, updated, and written, requiring two disk accesses. The stored vehicle image would have to be read to be used for matching, and then the new version would have to be written, requiring two more disk accesses. If a frame of photography contains 200 vehicles, this involves 800 disk accesses per frame, without taking into account the possible effects of "bad guesses" by the trajectory matching algorithms. This is not an excessive load on the I/O capability of a dedicated minicomputer. It would be quite expensive to tie up the same resources on a large time shared system, as well as probably being considerably more time-consuming.

The required disk capacity can also be estimated. A given vehicle image may be as large as that of a truck, with some margin around it - say 65ft x 12ft. With a ground resolution of 8 inches square per digitized value, this requires 1755 numbers to describe. If these are stored one per word (they could be packed two or, less conveniently, three per word), they occupy just less than one track on a standard disk pack. This in itself is convenient, since it means that each image can be stored as one physical disk record, simplifying the filing. A disk pack of the type used with the TRIM system (equivalent to the IBM 2311) has 10 surfaces with 200 tracks each. Thus one can without difficulty reserve 300 tracks for storing vehicle images.

Assume now that for the purpose of recording trajectories one assumes that a vehicle may be in the field of view of the aerial camera anywhere between one and five minutes (corresponding to average speeds for the mile traversed between 60 and 12 mph). In assigning record lengths for trajectories, one must assume that two values are to be recorded for every frame (i.e. once per second). Most probably these would be the distance along the road and some measure of lane position. A vehicle traveling at an average of 12 miles an hour would require 5 minutes to cross the field of view, so that a record some 600 words long has to be reserved for every trajectory to be encountered during a continous interval of observation. Two such records can be assigned per disk track. Assuming that 200 vehicles are visible at a time, and assuming now that they move at 60 mph -i.e. that 200 trajectories must be recorded for every minute of observation - one may deduce that the full disk pack (deducting the space used for storing vehicle images and miscellaneous storage) will accomadate records for about 15 minutes of observed traffic. In fact, since the record length and the number of records required were each estimated as worst cases and are mutually incompatible, one disk pack should accommodate the peripheral storage needs for an hour of observations, if reasonable assumptions are made about the rate of actual traffic flow in selecting record lengths. It is clear, however, that the total amount of storage capacity required is large.

The remaining major computer task that has not been discussed is the computation of the mapping between film coordinates and ground coordinates. In concept, what is involved is the computation of the camera location and orientation. This is based on the observed positions on the film of objects whose ground coordinates are precisely known. This initial calculation is rather complicated and must be precise. It is suggested that, if feasible, this task not be performed by the minicomputer, but instead be assigned to a larger machine with better arithmetic capability - in particular, one with longer word length and hardware floating point arithmetic. Such a division of duties is readily implemented if the minicomputer is equipped with an interface that allows it to communicate with remote teletypes via the telephone network. The resection program could then be implemented on an arbitrary large time sharing computer system, which would be programmed to accept data from a remote terminal and to return data to the same terminal via a telephone line. The minicomputer would need to be programmed to simulate the time-share terminal user, first sending character by character the control messages that an operator would give the time-sharing system, accepting and validating the systems acknowledgement messages, and then sending and receiving data in fixed pre-determined formats. Computer-to-computer data interchange based on such an approach is not inherently difficult to implement or expensive to maintain. The actual amount of computation on the time-share system would be relatively small. The main cost would be for maintaining the telephone connection - "terminal conect time" - which is billed at rates varying about $10 an hour ($6.50 - $20.00 for various system in the Cambridge area) by various time-sharing services, and telephone company charges, if any. It would certainly not be impossible to implement the resection routines on the minicomputer itself, but the alternative approach suggested here appears simpler and ultimately more efficient.

The remaining parts of the overall system merit little discussion here. The display oscilloscope and the tablet are for operator-computer system interaction. Their precise performance specifications are not critical. The display oscilloscope needs to be able to display both what is being scanned by the flying-spot scanner in a slaved mode and to display patterns according to the x and y coordinates and the intensity values commanded by the computer. In addition, it should have a scan generator (or share the scan generator of the scanner) to allow it to display in raster fashion an image stored in the computer as an array of intensity values. The scan generator should generate a raster with computer-selected horizontal and vertical limits and spacing, and accept the intensity values from the computer via a direct memory access channel.

The tablet may be any one of a variety of commercially available tablets that sense the location of the stylus on their surface and have a pressure switch at the end of the stylus, so that the computer can read stylus position and status (pressed down or not). The resolution on available tablets is 1000 or 2000 points in both the horizontal and vertical directions. This should be adequate.

4. SOFTWARE REQUIREMENTS

The software for a complete operational automatic scanning system involves a number of subsystems, which will be considered separately.

4.1 THE OPERATING AND FILE MANAGEMENT SYSTEM

This set of programs is absolutely specific to the particular computer chosen for the system. The greater part of the programs should be supplied by the computer manufacturer. This should include routines to communicate with the teletype and routines to read from the disk and write on it. The manufacturer should also supply an editor, assembler, FORTRAN compiler and other utility programs. The major addition to this set of programs that would have to be developed for an operational automatic scanning system is the disk file management routine. This routine would keep track of the physical disk record addresses of the data arrays describing the vehicle trajectories and the digitized vehicle images. Ordinarily, user programs store data on the disk and retrieve it by specifying logical addresses - file names and logically sequential positions within a file - to the operating system. The operating system maintains the file directories, pointers, record availability tables, etc., necessary to associate uniquely the logical file addresses with the physical disk addresses. The sheer number of different data arrays that must be selectively accessible to a scanning program (several hundred image arrays and ultimately several thousand trajectories) are likely to be beyond the capacity of the normal operating system of a minicomputer. Accordingly, either the operating program would have to be expanded for this specific purpose or a special user program would have to be written which would be allowed direct access to the disk and which would maintain its own version of a file directory. In following the first approach, one would have to take care to avoid or adjust for any incompatabilities that such an approach might introduce between the revised operating program and the various utility programs.

While perhaps aesthetically more attractive, it is probably the more difficult approach because it involves changing an existing complicated program, where any change involves the risk of introducing "bugs" difficult to diagnose when they manifest themselves later under some unforeseen set of circumstances. The second approach requires that large contiguous sections of the disk with known record addresses be reserved for the exclusive use of the special types of records. This is conveniently done by writing via the normal operating system large files which reserve the space. These files are thereafter not accessed by the operating system. Instead, a special program is written that essentially duplicates the operating system with respect to sections of the disk assigned to it. This approach is reasonably straightforward, and it leaves the original operating system and all associated programs, notably any diagnostic programs, intact and functionsl.

4.2 COORDINATE MAPPING ROUTINES

Generally speaking, the photograph of the highway to be scanned is taken from a helicopter whose precise position is not known a priori and changes from one frame to the next. The camera is aimed at some angle that is oblique to the ground and again varies with time. Thus in general, for each frame of photography the position and angular orientation of the camera must be established before the geometric parameters that allow one to compute the point on the film to which a point on the ground will project become available. The camera position itself must be computed for each frame from the position of known ground points within the frame. The computations involved are part of the subject matter of the science of photogrammetry.

In theory, if the focal length of the camera lens and the axial point on the film are known, six independent measurements are necessary to establish the basis for calculating the camera position and orientation. Six quantities determine the camera position - three rectangular coordinates describing (ideally) the position of the center of the lens and three angles describing the camera orientation. Theoretically a set of six simultaneous

non-linear equations could be written and solved for these six unknowns. As a practical matter, such a solution is likely to be excessively influenced by errors of measurement, both of the film coordinates and the ground control points. The approach used in practical cases is the following: A set of some 10-15 points on the ground are surveyed accurately and marked. The x and y film coordinates for each of these points are expressed in terms of their known ground coordinates and the unknown camera location and orientation parameters. A solution is then sought which minimizes the mean squared difference between the calculated and observed film coordinates. The equations take into account the geometry of projection through an ideal lens, modified by the distortion introduced by the actual lens. These are derived from a table based on measurements of the actual distortions introduced by the specific camera lens.

The various computations involved became quite complex. Fortunately, the equations have been developed, and algorithms for solving them have been implemented in FORTRAN and are available. In particular, computer routines to perform the mapping were developed at the University of California at Los Angeles in the course of the work there on aerial observation of traffic. The UCLA system of programs is based on a routine for computing the camera position developed by M. Keller and G.C. Tewinkel of the Coast and Geodetic Survey (Reference 1). This routine as originally given computed only the camera position and orientation. At UCLA, the additional routines to actually perform the coordinate mapping were added (based on the work of Prof. J. Anderson of the University of California at Berkeley), and some additional modifications to the basic Keller-Tewinkel program were made. The first of these was to increase accuracy, actually computing $\sin \theta$ where it was used, rather than using the small-angle approximation $\sin \theta \approx \theta$ of the original.

The second was to include an altitude interation loop for locating objects (vehicles) on the ground. The principle is the following: From a knowledge of camera position and orientation and coordinates of an object's image on the film, one can determine that the object itself lies at the intersection of the ground

surface and a line passing through the center of the lens (a known point) in a known direction. The horizontal coordinates x,y of the intersection depend on the height z of the ground surface at that point. If the highway is not perfectly level, the value of z can not be assumed to be a known constant. Instead, an approximate value is assumed, the corresponding x, y coordinates are calculated, and the actual height z of the surface at those ground coordinates is determined by a table look-up, with the table containing actual measured values. If the actual altitude of the point is sufficiently close to the assumed altitude, the ground coordinate values are assumed correct. If the actual altitude is different from the assumed, the actual altitude of the calculated point is assumed as the altitude of the desired point, and the process is repeated. With a reasonably smooth ground surface (as the highway must be), the process will convage rapidly (see Figure 2).

The third addition to the basic Keller-Tewinkel program made at UCLA is the addition of a final correction step to the calculation of the ground coordinates. To each coordinate pair (X,Y) computed, a correction term (ΔX, ΔY) is added, where

$$\Delta X = a_o + a_1(X-X_c) + a_2(X-X_c)^2 + a_3(X-X_c)^3 + a_4(Y-Y_c) + a_5(Y-Y_c)^2 + a_6(Y-Y_c)^3$$

and ΔY has a similar functional form.

X_c and Y_c are the ground coordinates of the center of the film frame and a_o-a_6 (and b_o-b_6 in the expression for ΔY) are obtained by a regression fit of the calculated values of the ground coordinates of observed points to the actual, known values of their ground coordinates.

There appears to be no theoretical basis for the form of the correction function, other than the observation that the average errors in the computed positions of known ground points would be reduced by its use. It is our judgment that the procedure used is unsound on theoretic grounds and should not be used. The basic fault of this technique appears to be that essentially it amounts to fitting functions with many adjustable parameters to a collection

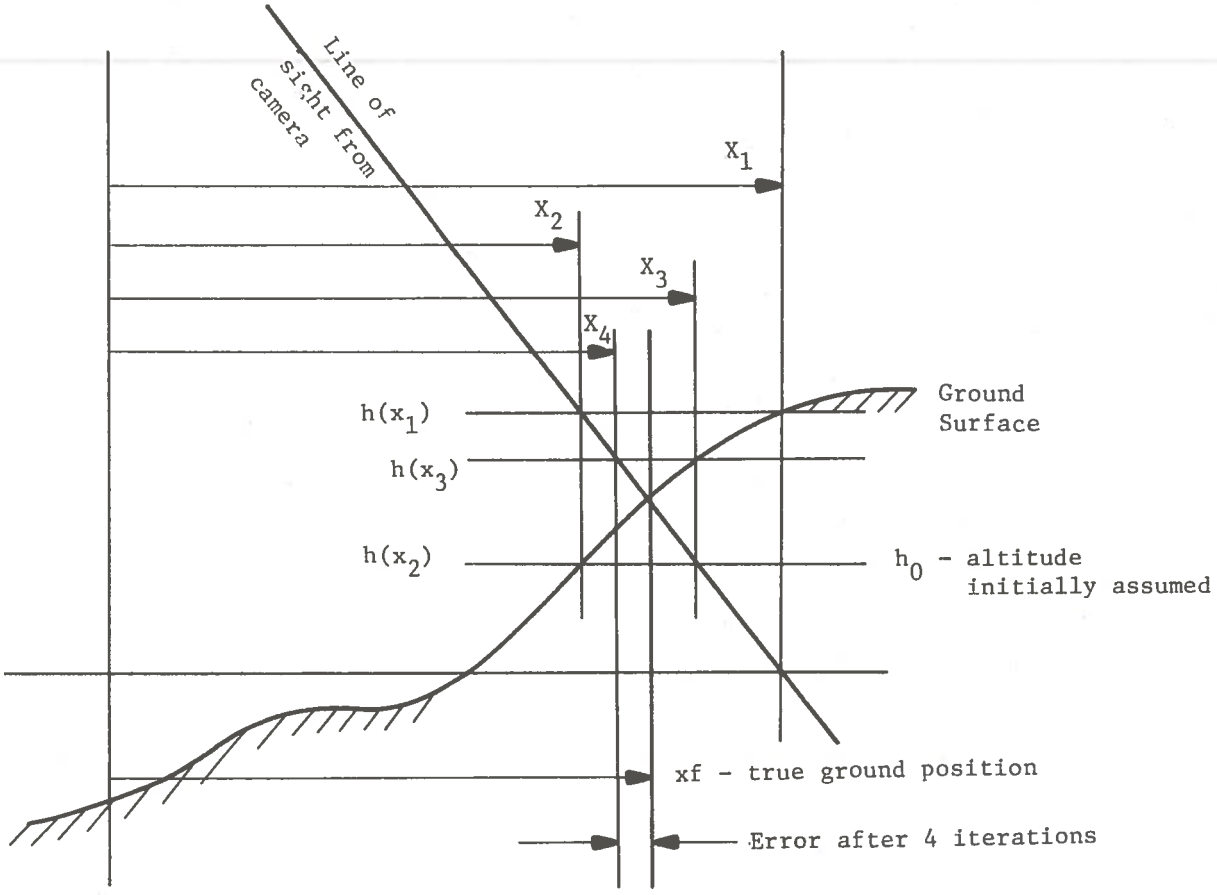


Figure 2. Steps in Correcting Error in Ground Position Due to Incorrectly Assumed Altitude. h (x_i) Is True, Known Altitude at Location $X_i.X_{i+1}$ Is Estimate of Position of Observed Point, Made Assuming that its Altitude Is $h(X_i)$

of relatively few sparse data points. A universal danger with such schemes is that the derived function will fit very well the points used in deriving it, but will assume rather widely varying values elsewhere, which may have no relationship to the underlying physical process. This danger is particularly acute when the functional form chosen is not inherently one that the data are known to satisfy. Thus the fact that the function chosen reduces the mapping errors at the known (surveyed) points does not per se mean that it will be helpful elsewhere. It may in fact be harmful. There is some evidence that this may in fact have occurred here; close examination of some computed vehicle trajectories showed the vehicles to be moving in parallel weaving paths. There was some suspicion, not verified, that this was the effect of adding the "correction" term.

It appears that the need may indeed exist for adding a correction term to the ground coordinates computed from film coordinates and camera position and altitude on the basis of a perfect (distortion-free) lens. However, these corrections should be based on the explicitly and systematically measured lens distortions. The effects of the imperfections of the camera lens, the optical system of the scanning apparatus, and the electronics of a flying spot scanner, if one is used, can be combined for this purpose.

It is suggested that the coordinate mapping programs be implemented on a large time-sharing computer. Although not essential, this is probably convenient. The program to be used could be the modified Keller-Tewinkel program discussed.

The location on film of the known reference points need to be given the resection program for every frame of film. Clearly, for the first frame they have to be specified completely with the ground coordinates for each point input via teletype or otherwise and then with the landmark being picked out on the display with the stylus. To locate the landmark precisely, the stylus should first be used to select a small area in the vicinity of the landmark to be scanned and to be displayed magnified, and then to pick the precise reference point on the magnified display.

For subsequent frames, the process can be simplified. The hovering helicopter should be sufficently stable so that the image of a given reference point in a film frame will be reasonably close to its position in the previous frame. Therefore, the system can be programmed in such a way that, when the film is advanced, the following sequence takes place for each reference point:

a. A small region of the film centered about the point where the reference point was located in the previous frame is scanned.
b. The scanned region is displayed magnified, and the operator selects the precise reference point on the display.
c. The computer associates the proper set of ground coordinates with this pair of scanner coordinates.

The reference points would be displayed one by one until all had been marked. Then the resection program would proceed to compute the mapping function. A special provision would have to be made for the possibility that some known ground reference points near the edges of the area covered might not be included in all photographs. This does not seem like a difficult problem to accomodate. Similarly, it is possible that, because of an unusually large camera movement, some or all of the reference points do not appear within the area scanned automatically. Again, a means must be provided for the operator to initate a rescan of the appropriate area.

It appears unlikely that successive pairs of photographs will generally be so similar that the alignment of reference points from frame to frame can be done fully automatically. This process would require that the small areas about the reference points from successive frames be brought into alignment by cross-correlation. The correlation peak would have to be found by exhaustive search, since experiments have shown that "hill-climbing" techniques are unreliable.

No statistics on the amount of camera movement between successive frames are available, and no detailed analysis of the required size of the areas to be matched, etc., has been made. However,

the following numbers appear reasonable and adequate for estimating the time that an automatic routine might take to find a reference point with no a priori knowledge except its location in the previous frame. Assume the scanner element to represent an area 9 inches square on the ground. Assume that the camera movement is such that the ground point imaging at a given film position moves no more than 20 feet in the interval between frames. Then a circle of radius 20 ft or 27 resolution elements needs to be searched for the match. This gives about 2300 points which have to be examined as candidates. At each such point, the cross-correlation between the area about that point and the area about the actual reference point as observed in the previous frame must be computed. Assume the incremental areas to be matched to be about 22 feet square. Then each contains 900 resolution elements. Assuming 25 μsec calculation time per pair of elements, one would need 22,500 μsec per point considered. Given over 5000 points to consider, it would require more than 45 seconds to establish the best match. This is unacceptably long, and shows that the approach described above, requiring the operator to indicate the reference points with a stylus, is required.

The question is left open whether a combination of the two approaches might not be efficient. The number of ground reference points measured is considerably greater than that theoretically necessary to compute the mapping function. The redundant points serve to reduce overall errors. It may be practical to have the operator indicate the locations of only enough reference points to allow the mapping to be computed coarsely. It will then be possible to calculate quite good approximate locations for the remaining reference points. The precise location of these can then be established automatically by area correlation techniques (with a considerably smaller region to be searched for the best fit then that considered above), and the final mapping function can then be computed based on all available reference points.

The efficiency of such a hybrid scheme is somewhat difficult to predict in the abstract. Such a scheme should be considered as a refinement to be added to the system dependent on operator input of all reference points, to be evaluated with the actual hardware

available for the final system. The programming effort required to implement this approach should be made, since the various subroutines required for this will have to be provided for other tasks in any case.

A final feature should be incorporated in the system to prevent major errors due to operator blunders. The positions of the reference points as measured should be checked against their calculated positions. Excessive discrepancies (errors of more than some 3-4 feet) should be taken as evidence of input errors, and some correction stage should be entered.

4.3 IMAGE MATCHING PROGRAMS

The image matching routines are central to the automation of the data reduction process. The success of the concept of automation depends on the speed and accuracy with which the matching can be achieved. It will be evident that the speed of the process is inversely proportional to the area which needs to be searched to locate a match. The accuracy of the vehicle location process must be defined in terms of two figures of merit: The probability that a vehicle observed at a given instant is correctly associated with its trajectory (consisting of the records of its positions at previous instants of observation), and the positional accuracy with which it is located - i.e., some statistical measure of the distribution of deviations between its position as recorded and its true position on the ground. Clearly the first is the more important; positional accuracy in measuring the location of the wrong vehicle has little merit by itself. The probability of matching trajectories correctly is a function of the trajectory matching routines (see section 4.4) and the basic predictability of vehicle movement. It affects the accuracy of location measurements only in that if the position of a vehicle is not predicted with reasonable accuracy a priori, the area that needs to be searched for the vehicle position is large, and the probability arises that either of two situations may occur, which result in an inaccurate determination of position. (a) In an effort to reduce the time of searching for the vehicle, the area to be searched is made so small

that the vehicle in fact is not contained in it, so that its location can be determined. (b) The area to be searched is so large that another vehicle is within it, and it confuses the search routine.

If these potential problems are ignored for the time being, then the problem of measuring vehicle location can be formulated as follows: There is a given two-dimensional array of measurements and a smaller array, which is assumed to be a segment of the larger array after distortion by various noise processes. The problem is to determine which of the segments of the large array the small array corresponds to. This is a problem of map matching with applications in other fields. It has been analyzed and described in some detail by M. A. Fischler in a Lockheed Missiles and Space Company Report (Reference 2). Fischler formulates the problem as one of matching a "sampled map" of N measurements against a set of "reference maps" of the same size, where each reference map is a subset of the large map.

The nomenclature used by Fischler is not descriptive of the role that the various maps play in the current problem, but the model is applicable. The role of the "sensed map" is assigned to the image of the given vehicle in the previous photograph with a border of surrounding road surface. The set of "reference maps" are the various subareas of the same size within the current frame against which this image is to be matched. (We make the plausible assumption that the pavement forming the background for the vehicle is sufficiently uniform, so that two photographs of the vehicle on different pieces of pavement are essentially indistinguishable if the vehicle is in the same position relative to the edges of the photograph.) Ignoring now the possible effects of rotation (of the camera or because the vehicle has turned), of scale (camera altitude changes) and of "rubber sheet" distortions of the image, the given problem reduces to the simplest case considered by Fischler - a sampled map that is a copy of one of the reference maps distorted by additive noise. The "sampled map" may be described as a two dimensional array of values $S_{x,y}$, $(1 \leq x \leq N_x,\ 1 \leq y \leq N_y)$. The set of

reference maps may be described as a set of such arrays $r_{x+a,y+b}$ ($1 \leq x \leq N_x$, $1 \leq y \leq N_y$) where each a,b pair characterizes one reference map of the set. (Geometrically, a and b represent the distance in the x and y directions of the particular reference map from the coordinate origin of the larger map from which it is drawn). Fischler shows that the best alignment of the sampled map with a reference map is that for which

$M(a,b)=\sum\sum_{xy}(S_{x,y}r_{x+a,y+b} - 1/2\ r^2_{x+a,y+b})$ is the minimum over the set of a,b pairs considered. In the current case, one may assume that all the reference maps considered contain a vehicle image surrounded by empty pavement. The quantity $\sum\sum_{xy} r^2_{x+a,y+b}$ will then be essentially the same for all the maps considered, since presumably the various elements of pavement near the edges contributing to the sum will always contribute similar amounts. Then the best fit criterion can be simplified to be that the best match is that for which $R(a,b) = \sum\sum_{xy} S_{x,y}\ r_{x+a,y+b}$ is the maximum. The quantity R(a,b) may be recognized to be the two dimensional cross-correlation of the stored vehicle image from the previous frame and the segment of road scanned in the current frame. The point of best match is the correlation peak.

We may calculate the time required to perform the search of a given area as follows: The double summation encompasses all measurements over the area covered by the vehicle and the surrounding pavement. Assuming an area roughly 25ftx10ft is scanned, and the resolution of the scanner amounts to 9 inches on the ground, there are a total of some 450 measurements per image. The computation of the double sum of products requires about 25 machine cycles of 1 μsec each for the calculation of each term and the attendant indexing, etc. Thus, the calculation of each value of the cross-correlation requires about 11 milliseconds (assuming all required data are already in core memory). If the position of the vehicle is assumed to be known to within 12ft along the direction of travel and within 4ft laterally, then a cross-correlation function spanning a rectangle 24ft x 8ft must be computed. On a nine-inch grid, this requires approximately

340 cross-correlations to be computed. Thus the total search time requires about 3.5 seconds per vehicle, not counting the time consumed by data acquisition, coordinate mapping, disk file accessing, etc. Assuming these to be negligible, we must still note that the rate of 1 vehicle/3.5 seconds is not significantly different from the 10 vehicle/minute rate achieved by manual methods at UCLA - a rate considered too slow.

Several means may be employed to increase the speed. If the cross-correlation function is computed at every other point in both the x and y directions to obtain the approximate location of the maximum, and then computed at every point only in the immediate vicinity of the maximum to get its exact location, a nearly four-fold improvement in speed can be achieved. Similarly, one might use only every other point in the picture arrays ("maps") themselves in the rough computation of the cross-correlation again to achieve a nearly four-fold improvement in speed. These measures in combination could reduce the total search period for a vehicle to less than half a second. This would represent an improvement of an order of magnitude over the manual system and be a justification of the automated approach, assuming the area to be searched for the match can in fact be kept as small as indicated.

The accuracy of the vehicle position determined by this means depends on the sharpness of the correlation peak. Inevitably, random noise will be present. Its effect will be to cause the maximum value of $R(a,b)$ to be achieved for some pair (a,b) other than the proper one whenever the additive noise makes the noisy correlation function assume its largest value at the wrong point. Clearly the probability of this happening decreases as the differences between the noise-free correlation peak and the other values of the correlation function increase with respect to the value of the noise terms.

The nominal factors that determine the sharpness of the correlation peak are image contrast and image detail. The role of contrast is obvious - it is equivalent to signal amplitude. The finer the detail in the two images being correlated, the more quickly the correlation drops as the image details are taken out of alignment. This implies, for instance, that automobiles with roof-tops and hood and rear decks in contrasting colors could be more accurately located then delivery vans in one color, that distinct and relatively short shadows cast by a vehicle would be helpful in locating it, etc.

The features of images that make it possible to align them accurately are their edges. This consideration leads one to consider the possibility of performing spatial filtering of the images in such a manner as to enhance the edges in the hope of sharpening the correlation peaks. The idea appears to have merit when one examines the idealized case (Figure 3). The auto-correlation of a one-dimensional square pulse is a triangular function, and clearly the effects of noise could have the effect of shifting the apparent maximum. If the pulse is subjected to filtering by a simple one-demensional "discrete Laplacian" filter commonly used for edge enhancement, the resulting function is a series of short bipolar pulses whose auto-correlation displays various small oscillations, but more important, has a very sharply defined maximum peak that appears unlikely to be obscured by uncorrelated additive noise.

In practice, the apparent benefits of such spatial filtering turn out to be somewhat illusory when applied to the present problem. The actual results achieved in computing cross-correlation of vehicle images subjected to spatial filtering of the class described have the appearance of high-frequency noise.

The reasons appear to be these: Vehicle images consist largely of areas of uniform shade. After filtering, the only contributions of "signal power" came from the edges themselves, and represent a relatively small fraction of the "power" of the original signal. The "noise power" of the additive random

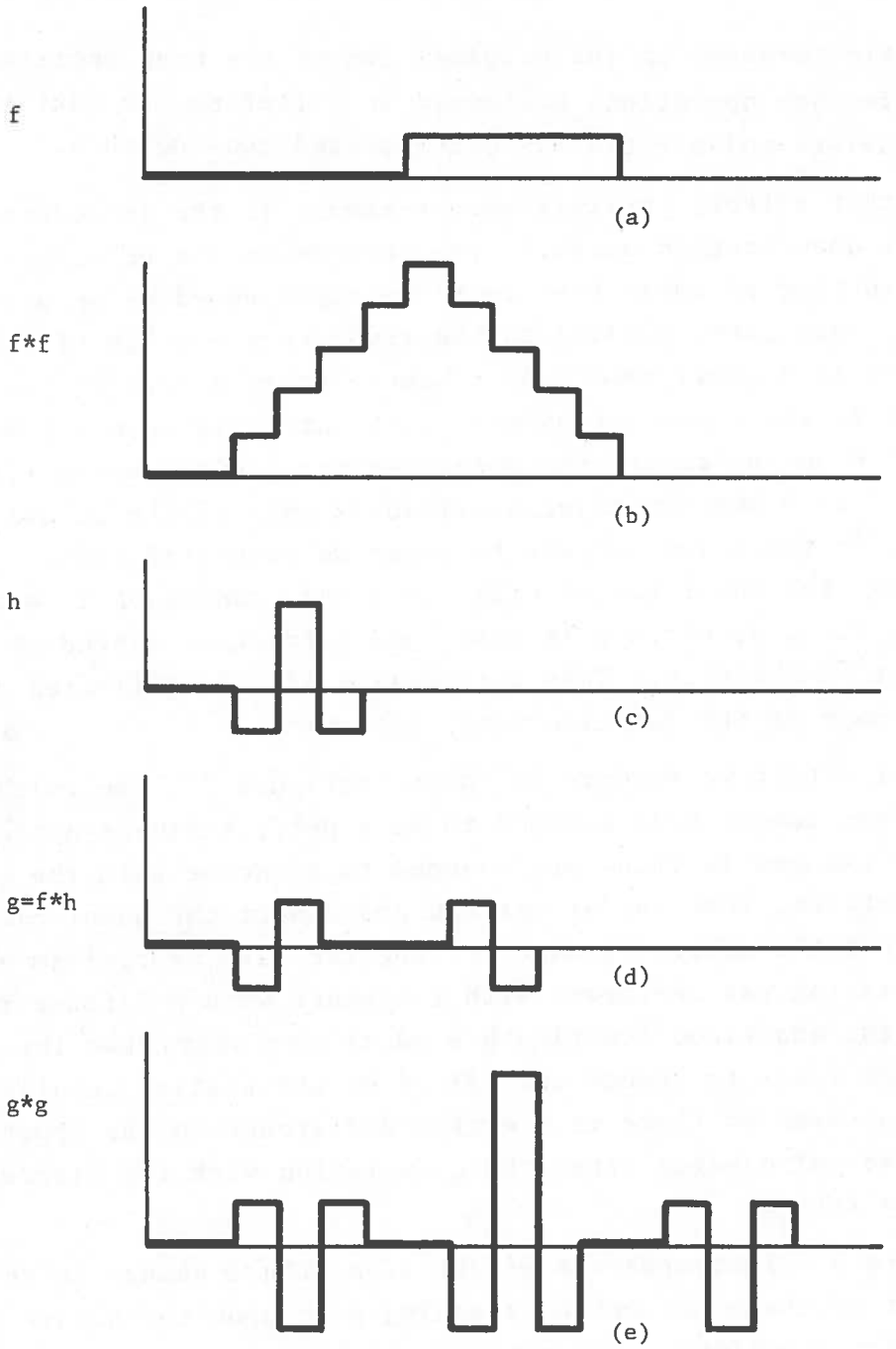


Figure 3. Improvement of the Sharpness of the Auto-correlation Peak by "Discrete Laplacian" Filtering of Idealizer Pulse Waveform

fluctuation present in the original images has been amplified by the difference operations performed in filtering, so that the overall signal-to-noise ratio has deteriorated considerably.

Another effect, possibly more serious, is the introduction of "spatial quantization noise." The phenomenon can be understood by considering an image that is to be represented by an array of numbers, each corresponding to the true average value of the image intensity in a small square of a superimposed grid. If now there is in the image a boundary, with intensity values 1 on one side and 0 on the other, the quantized array will show a transition of 1 to 0 between adjacent elements only if the actual boundary in the image happens to coincide with grid line. Otherwise, the quantization will cause the transition to appear as 1, 1, p, 0, 0, where p is some random fraction dependent on the placement of the grid. This variability will be reflected also in the shape of the function after filtering.

An illustrative example is shown in Figure 4. The original unquantized function is assumed to be a perfect square pulse, but the step changes in value are assumed to coincide with the edge of a quantizing interval at one end and bisect the quantizing interval at the other. Assume for the the sake of realism that the quantizing was performed with a scanner with a diffuse spot, so that the quantized function has edges less sharp than the original. This tends to reduce the effect of the spatial quantization noise, but even so there is a marked difference in the appearance of the two pulse edges after the convolution with the discrete Laplacian filter.

Figure 5 illustrates the effect of a slight change in the alignment of the pulse and the scanning grid upon the nature of the correlation functions. The auto-correlation function obtained by convolving the filtered signal with itself shows a single distinctive sharp peak at the point of coincidence. In contrast to this, the cross-correlation function of the filtered wave forms due to the two slightly shifted pulses shows a much lower peak at the point of best alignment of the two original pulses, and

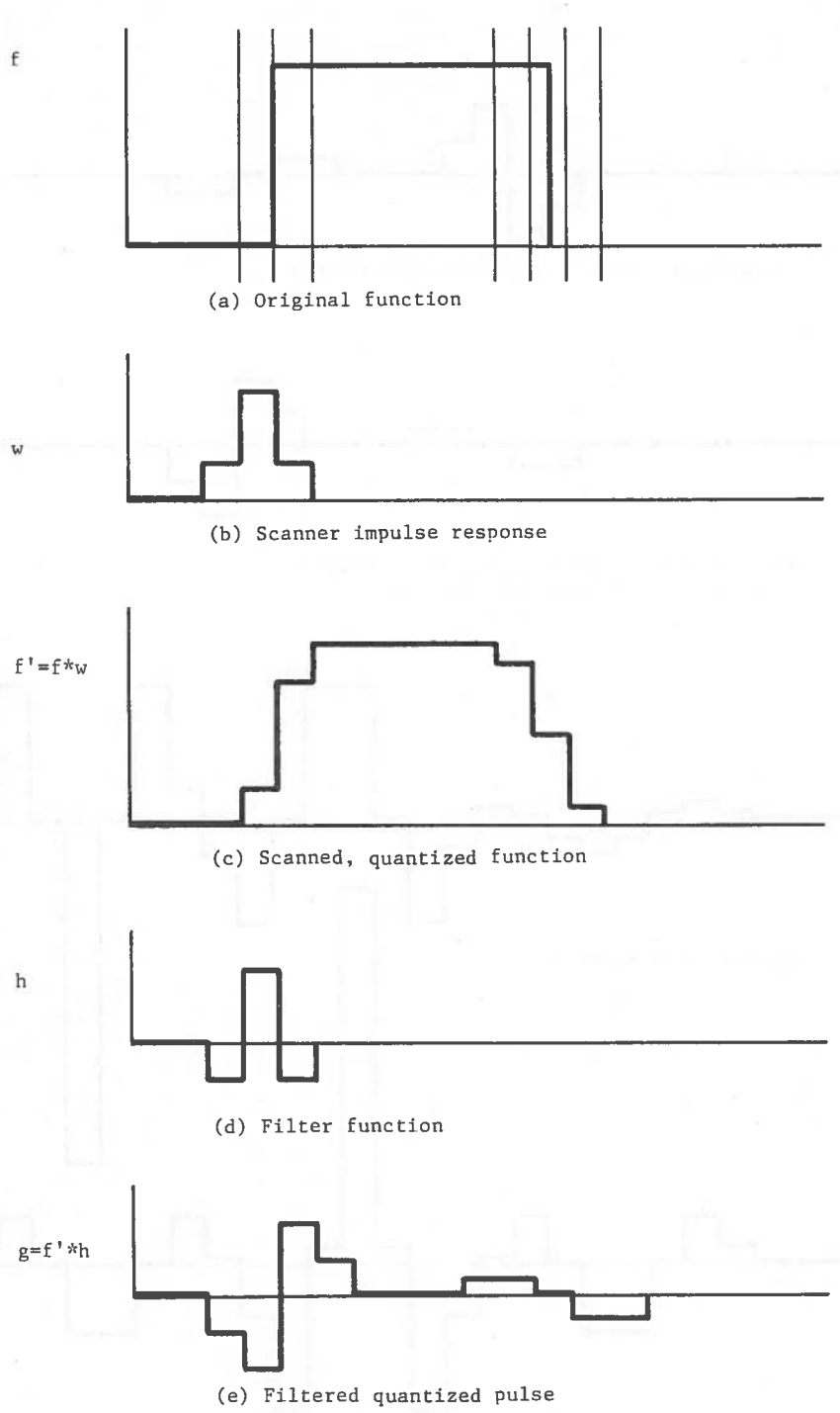
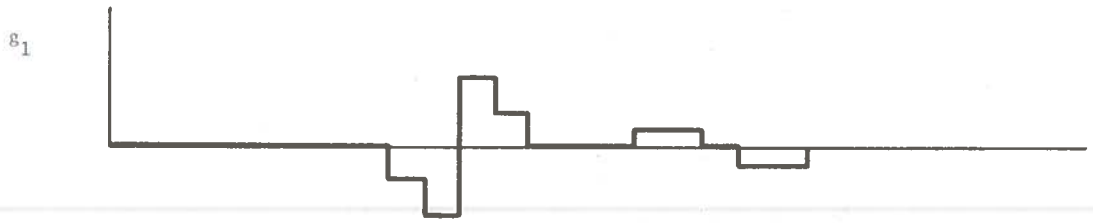
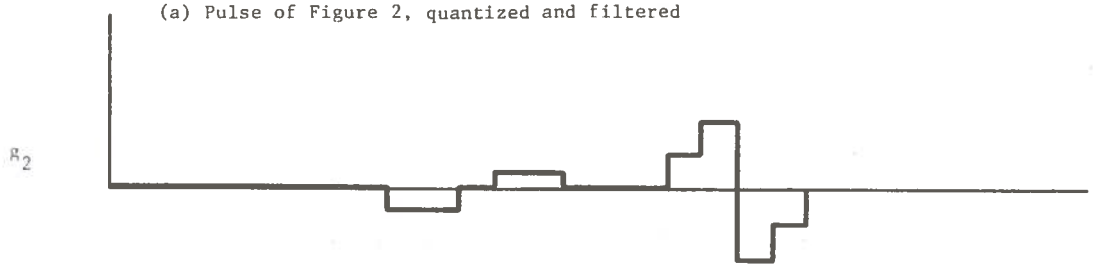


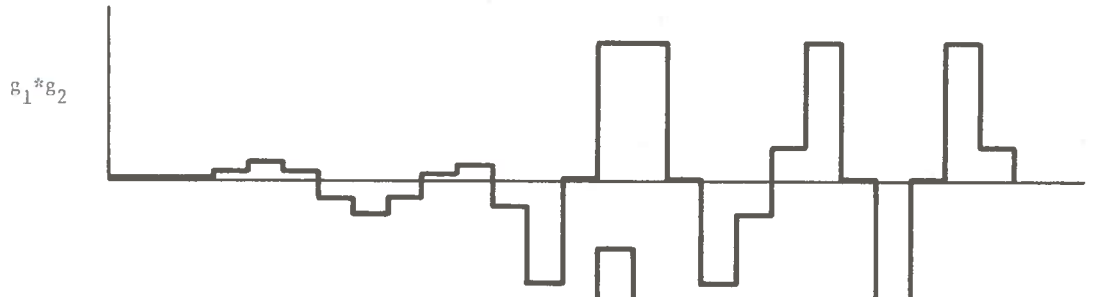
Figure 4. Effects of Spatial Quantization



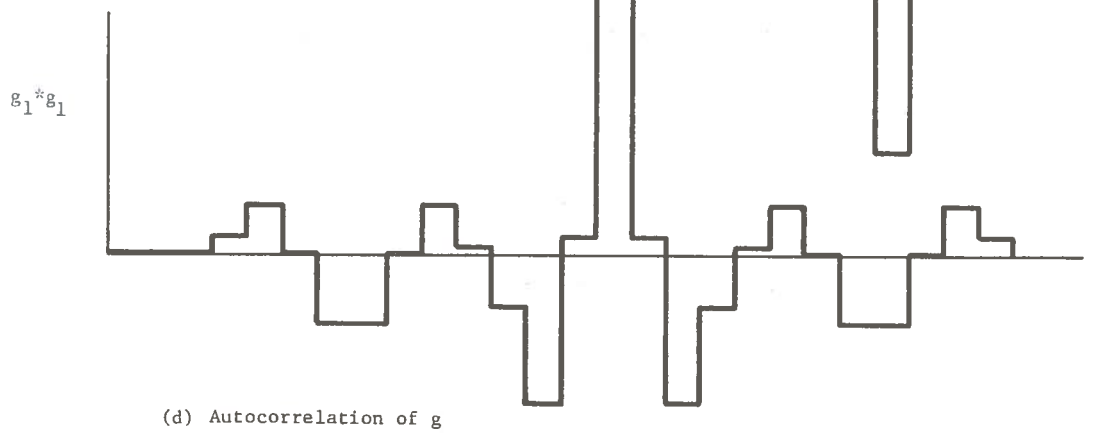
(a) Pulse of Figure 2, quantized and filtered



(b) Same pulse, shifted half a grid increment to the right, quantized and filtered



(c) Crosscorrelation $g_1 * g_2$



(d) Autocorrelation of g

Figure 5. Effects of Spatial Quantization Noise

what is worse, shows two other equally high peaks far away from the point of desired best match, making the function totally unsuitable for the proposed purpose.

Since the correlation peak of the basic quantized images without filtering is quite flat, the point of best coincidence can not be defined very reliably. Experiments on the TRIM system showed correlation peak areas some 5 or 6 grid positions across, within which the variation from point to point was quite small, so that the maximum within it was likely to be determined by noise. The images considered were distinct (i.e., separately scanned) versions of the same vehicles at the same locations. Differences due to comparing different photographs would make the results even worse.

These results indicate that it is important to minimize the additive noise. To this end, in experiments with the TRIM system images were scanned repeatedly, and the average reading was used. As a practical matter this means that the scanning must be performed more slowly. The computations involving scanner speed and those involving the rate at which correlation functions can be computed show that this would not lead to a significant decrease in overall processing speed.

It is likely that some spatial filtering can be performed on the scanned images to improve the reliability with which the correlation peaks can be assumed to occur at the point of best alignment. Such filters would have the effect of perfoming some edge sharpening (i.e., some high-frequency enhancement) without blocking the low frequency components of the images (as the "discrete Laplacian" filter illustrated does). The precise nature of the filter would have to be determined using the actual eqipment being used and representative copies of the film being scanned, since the shape of the best filter impulse response will depend on the noise properties of the scanner, the signal strength (i.e., contrast) of the film, and the resolution of the scanner. (If the scanning spot is larger than the nominal size of the grid increments being scanned - i.e., there are more addressable points

than physically resolved points in the image – resolution suffers, but the spatial quantization noise drops.)

It is considered unlikely that substantial improvement will be achieved in sharpening the peaks of the correlation. If the "region of uncertainty" remains some 5 scanner grid elements across, then with a scanning grid assumed to give 9-inch resolution on the ground, a "region of uncertainty" some 4 feet across can be assumed, which means that vehicle location can be read to within $\pm$ 2 or 3 feet. This is comparable to the accuracy achieved by human film readers.

The discussion above has assumed that the appearance and the orientation of the vehicles will be constant from frame to frame, so that the matching algorithms can operate on what are presumed to be identical images distorted by additive noise. In fact neither of these assumptions is entirely valid. Clearly a vehicle directly under the camera will appear in the photograph only as the vehicle top and as any shadow cast by the vehicle. That same vehicle, when it has moved to the edge of the scene, will show some of the essentially vertical surface of its rear end, and it would obscure part of the shadow that it might be casting ahead. The changes of this nature, however, are minor between successive frames and can be safely ignored for any pair of adjacent frames. They do not accumulate, because our concept of system operation involves storing the current view of each vehicle for matching against the next frame only. This same approach also eliminates any practical problems that might come about as a result of camera altitude changes and resultant scale changes.

A class of changes in the vehicle images that is of practical importance is that introduced by vehicles turning, and to some extent by rotations of the camera about its own axis. The nature of the problem is easy to illustrate. If one overlays two identical square grids so that they coincide exactly, and then even slightly rotates one, he finds that the individual grid squares that originally overlapped completely very quickly get out of alignment and soon do not even partially overlap. The effective-

ness of any image matching scheme that is based on a grid increment by grid increment comparison of two images is certain to suffer as a result.

Fischler (op.cit.) treats the case in terms of statistical decision theory and proposes in effect the following: In computing the decision function analogous to the cross-correlation used for the simpler case above, the effect of the grid elements is to be weighted by a function of the angle of rotation and the distance of the elements from the center of the grid. The elements near the edges eventually get completely out of alignment and contribute nothing to the decision function. For a rotation of four degrees, no grid elements outside a 30 x 30 square would contribute (and the elements near the edges would contribute very little).

The decision function is relatively complex, both to derive and to calculate. An attempt to apply it would require several times as much calculation for every candidate point, making the process significantly slower. Furthermore, since the most significant information about vehicle location is contained implicitly in the location of the vehicle sides – which would fall nearer the edge of the grid array, and by this rule would be given relatively little weight – the technique is unlikely to work very well.

An alternative method was devised at TSC to attempt to overcome the limitations inherent in a fixed grid. In principle the method is still to try to match the rotated grids square by square, but to choose as the pairs of squares to be matched not the pairs of squares in the two grids having the same nominal coordinates as before, but rather the ones most nearly coinciding physically. The idea is implemented by constructing approximations to square grids oblique to the grid imposed by the hardware. The measurements corresponding to any raster scan line are taken to be the measurements from the sequence of points along the actual grid that lie closest to the desired oblique scanning line (see Figure 6). By this means, images slightly oblique to each other can be brought into reasonably good alignment and the best match can be sought by cross-correlation as before. The algorithm devised for the oblique

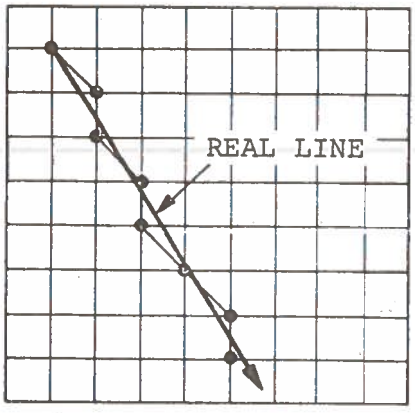


Figure 6. Approximate Oblique Scan

scanning, which essentially amounts to rearranging the order in which the digitized image array is stored in core, is quite fast, taking some 22 μsec per pair of coordinates. Thus, it would not involve an unacceptable increase in the time for a search (roughly 10 milliseconds per vehicle, if the vehicle image from the previous frame, being the smaller array, is rotated into alignment with the presumed current orientation). The approach does not fully solve the problem of misregistration, for there simply are not pairs of squares in rotated grids that coincide exactly. Figure 7 shows the pattern of displacement of scanning element centers from their desired locations when a grid roughly the size of a car at the resolution we have considered is turned five degrees.

A further possible approach to the problem would be to attempt to simulate the results of scanning a raster oblique to the actual raster by interpolating between the values of the surrounding known elements. This has not been attempted, since it would be relatively time-consuming and would require a different set of weights for the interpolation function at every point of the array,

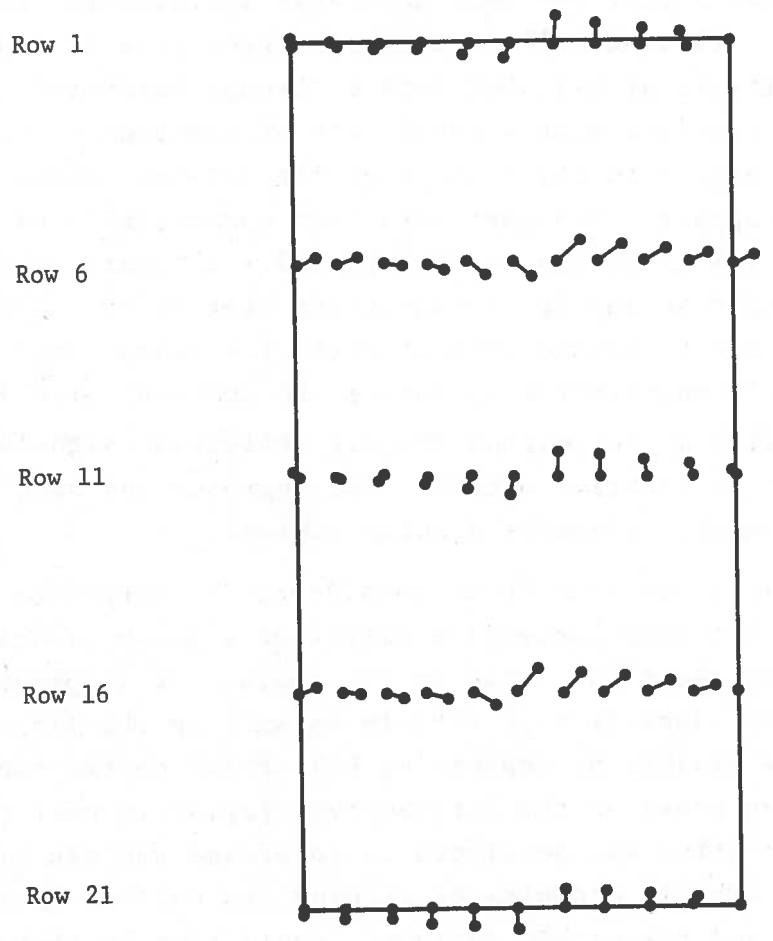


Figure 7. Misalignment Due to Oblique Scanning — A 0° Vehicle Vs. a 5° Vehicle

since every assumed oblique grid point has a different set of distances to the actual grid points surrounding it. The process of interpolation in any case amounts to taking a weighted average of the surrounding values and thus to a blurring of detail. As such it would tend to blur the correlation peak as well.

It appears that the most desirable approach for solving the problem of rotational discrepancies between vehicle images is not through software at all, but rather through hardware. If the scanner is provided with a capability of changing the scanning axes with respect to the film, then the problems described largely disappear. Scanners have been commercially marketed in which the film drive can be rotated under automatic control with the scanning tube and deflection yokes stationary. The benefits of such a feature should make it worth the added complexity and cost. An alternative way to rotate the scanning grid is electronically by subjecting the x,y deflection signals to multiplication by a rotation matrix. The approach has been used in building computer graphics display devices.

The final function to be considered in connection with the process of matching successive images of a given vehicle is that of acquiring the first image in the chain. It is proposed that the operator identify each vehicle as such in the first frame in which it is visible by depressing his stylus on the tablet when the tracking cross on the interactive display is over the vehicle image. A routine was developed to determine vehicle size and define the area it occupies by tracing its outline in its image. That area, and reasonable surround, could then be stored in the reference image for matching in the next frame.

The idea as such appears to be valid, and the contour tracing routine worked. However, since it works by finding the boundary between areas lighter and darker than some arbitrary thereshold, it was found to be quite sensitive to the setting of the threshold value. Using the TRIM system, it was found that the threshold frequently had to be reset by trial and error when the system had been turned off, and that the same threshold did not even always

work satisfactorily for different vehicles in a single frame of photography.

On the TRIM prototype system, the success of the outlining operation could be observed by seeing the contour obtained displayed as a bright line over the display of the vehicle itself. This capability, combined with a convenient means of altering the threshold (such as a thumbwheel whose position the computer could sense), should allow the operator to monitor the system operation and correct any potential errors immediately.

The shape of the area surrounding the vehicle should be a rectangle aligned with the edges of the road. It has been implicitly assumed in the discussion so far that the road would initially be parallel to a scanner axis, so that such a rectangle could be scanned by normal raster scanning. Any other conditions introduce complications. Using the "oblique scanning" algorithm is not desirable, since it is slower and must be done strictly by computer selection of coordinate pairs (rather than by an automatic raster generator).

Furthermore, if an attempt is made to match point by point two "obliquely scanned" grids that have been rotated relative to each other, the actual distances between supposedly corresponding points turn out to be significant. The effect can be observed by comparing Figures 7 and 8. The misalignments between supposedly corresponding points of two "obliquely scanned" grids, one rotated 10° and one 15° from the basic raster, are larger than the misalignments between the actual grid and an "oblique scan" approximation of a grid rotated 5° from it. One can not use as the reference image of a vehicle a rectangular array whose sides are parallel to the grid axes but not approximately squarely aligned with the road. An area sufficiently large to enclose the desired vehicle would have corners projecting into adjacent lanes, and attempts to match vehicle images would be confused by the appearance of other vehicles in adjacent lanes.

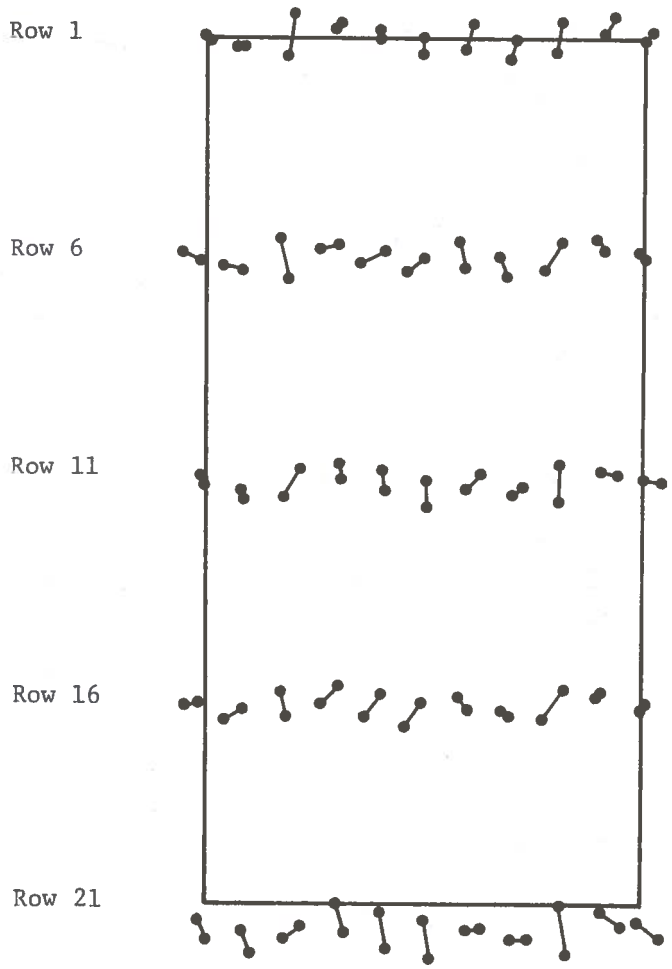


Figure 8. Misalignments Due to Oblique Scanning — A 10° Vehicle Vs. a 15° Vehicle

Theoretically, one could select as measurement sets to cross-correlate areas of square raster scan arrays with boundaries oblique to the "rows" and "columns" of the array. The difficulty with this approach is that it would complicate the indexing calculations during the calculation of the cross-correlation functions, with a significant increase in computing time. From the foregoing, it appears again that the most desirable solution is to eliminate the problem by including in the system hardware the capability to select the direction of the scanning grid axes.

4.4 TRAJECTORY MATCHING ROUTINES

At the time this project of investigating the automation of the film reading task was begun, it was assumed that the routines for vehicle trajectory computation would be available from work done elsewhere. In particular, such routines were being designed in conjunction with the manual film reading efforts underway at UCLA and SDC under FHWA sponsorship, and it was assumed that these could be applied, with minor modifications, to the task at hand.

It has been concluded since then that the routines available would not be adequate for the purpose. No significant work on this problem has been performed at TSC. However, because successful performance of this step is absolutely essential if automatic film reading is to be performed, the problem will be discussed here.

In a purely manual film reading system, the trajectory construction problem is essentially this: There are given a set of vehicle positions for each of a number of vehicles at successive instants of time, but the vehicles occupying the positions are not identified. One must construct the trajectory of each vehicle – that is, group together those sets of vehicle positions that are the successive locations of one specific vehicle - strictly by reasoning based on the constraints that vehicles move in continuous paths with limited capabilities to accelerate, decelerate, and change lanes and without impinging on the same space at the same time.

As a practical matter, such matching reduces to considering succeeding instants of time (photographs) one by one, and associating one of the observed vehicle locations with each of the trajectories to be extended. In a manual system, it is convenient to perform this extension by extrapolating the available trajectory to determine the limits of the area in which the vehicle might be, and then to assign as the trajectory extension that observed vehicle location which falls in this area (and which is in some sense best if there is a choice of several possibilities).

Essentially, the same process must be used for the automatic system, but the requirements are more stringent if the system is to be practical. The extrapolated area of possible locations must be reasonably small in extent. Otherwise, the search for a match is too time consuming, and no benefits are achieved by automation.

The set of SDC-UCLA programs as presently written are not good enough for the automatic system; the size of the region predicted as possible for a vehicle in the traffic stream is too large to be searched in an acceptably short time. Thus, the position of a vehicle assumed to be traveling 60 m.p.h. is predicted one second later to be written $\pm$ 29 feet, requiring that an area 58 feet long be searched for a match. Lateral movement in the course of one second is assumed to be as much as one lane width. These limits must be sharply reduced (at least on the average) to allow reasonable search times. Unfortunately, statistics apparently have not been gathered that would allow reliable computation of the size of the prediction region that is made necessary by the inherent instabilities in the traffic flow itself.

In trajectory matching using manually gathered vehicle positions, the size of the extrapolated regions for each trajectory is not per se a critical parameter, since it does not influence the data collection. At worst, excessively large regions should complicate the task only in that they should lead to occurrences that several candidate vehicles appear as possible extensions of

the trajectory. By definition, a good trajectory matching routine should sort these out properly. Unfortunately, the UCLA-designed trajectory construction routines do not give satisfactory results, the percentage of trajectories carried currently through the test section apparently being as low as 50% for some time intervals and never better than 90%. Very briefly, the routines work as follows: Every vehicle in a given frame of photography is located and assigned an estimated velocity. The next frame is then scanned in its entirety, and every vehicle in it is located. Then the attempt is made to identify each vehicle from the first frame with the vehicles in the second. The following steps are performed for each vehicle in the first frame (the vehicle order depending on the order in which they were scanned): A rectangle is constructed which specifies the limits of the area within which the vehicle is constrained to be in the next frame (based on its physical acceleration and deceleration potential and the vehicles around it). Its most likely position in the next frame is computed (based on current position and estimated velocity). The list of vehicle positions in the next frame is now searched for all vehicles that fall into the rectangle of possibilities and that have not already been identified with some vehicle in the first frame. The one closest to the extrapolated "most likely position" is assumed to be the vehicle sought, and is tagged to remove it from consideration in constructing succeeding trajectories. If no previously unassigned vehicle is found within the limits of the rectangle of possible positions, the operator is instructed to rescan the general area in the expectation that the vehicle sought might have been inadvertently missed in scanning the frame originally.

One of the chief causes of mis-matched vehicle trajectories is lane-changing. Each vehicle is assigned a potential of changing lanes to its right or its left, depending on the configuration of vehicles around it and their speeds. This algorithm is quite crude and apparently has not been tested for conformity with

real driver behavior. In addition, there is some reason to believe that the program implementing it may contain an error. As described, the algorithm never allows a trajectory to be constructed that includes a lateral vehicle movement of more than one lane in the interval between sucessive frames. However, one illustration of a mis-matched trajectory found by SDC shows an incorrectly constructed trajectory that involves a lane change of two lanes occuring in a one-frame (1 sec) interval. The illustration is presumably of a typical case and was described as being an example taken from the actual data.

The other main cause of errors in matching trajectories apparently is that the initial assumption of vehicle velocities for vehicles first entering the study region tends to be bad. This is computed as a function of lane occupancy only, ignoring for instance such things as the distance to the vehicle immediately ahead and that vehicle's velocity. In the UCLA system the operator locates vehicles and the computer attempts to extend the trajectories while the film is still in position and the operator can confirm vehicle positions by re-reading sections of the film. If, in attempting to extend a trajectory, no candidate vehicle position is available, the operator is then instructed to rescan the five vehicles nearest the position where the vehicle ought to be. The hope is that a vehicle inadverently omitted will be picked up.

There are what appear to be weaknesses in the concept. The most serious is that not all the rescanned points are tested to see which is most likely to be an extension of the trajectory in question. Instead, the first possible one that falls inside the allowable limits is immediately taken.

After all the trajectories for which it has been possible to find continuations have been extended, all scanned points that have not been matched to a trajectory are considered to be spurious and are eliminated from the table. Thus, once a vehicle becomes disassociated from a trajectory, it will always thereafter be treated as a spurious point and dropped from consideration.

Not only is its own trajectury then lost, it is not properly taken into account in computing the possible extensions of trajectories of the vehicles around it, thus quite likely also causing other errors.

The basic problem with the algorithm for constructing trajectories from vehicle locations identified by a human operator seems to be that, even with all the data available, too limited a set of considerations is used in selecting the point to add to a particular trajectory. The problem is difficult to describe concisely, but is basically a matter of neglecting the information inherent in the locations and presumed movements of surrounding vehicles in attempting to deduce the movement of a given vehicle about whose identity there is some doubt. The following is an attempted explanation: Consider a set of trajectories labeled A,B,C,D,etc., which have been constructed to include everything scanned as far as the n-th frame. We have now scanned all the points representing vehicles in the n+1st frame, and have stored these in a table, where they are labeled a,b,c,etc. The task now is to attach them to the proper trajectories. The current algorithm proceeds by trying to extend the trajectories one by one. Thus, it would compute a most likely continuation point for trajectory A by extrapolation. It would then search exhaustively through the list of observed locations a,b,c,d... for the one closest to this extrapolated point. Assume this to be point b. A test would be performed as to whether point b fell into the area defining the possible continuation of trajectory A. If the test were satisfied, point b would be attached to trajectory A and removed from the list of possible continuations for the remaining trajectories. An attempt would then be made to extend trajectory B in the same way, considering as candidate extensions points a, c, d, etc. Point a might now be found to be the closest to the extrapolated position, but outside the allowed area. Thus it would not be tied to trajectory B. Instead, trajectory B would be tagged as unextended and a rescan would be requested in the vicinity of its possible extension. The rescan would not be performed until all other trajectories had been extended,

insofar as this was possible. Assume that all other trajectories are completed satisfactorily. In rescanning, the coordinate values of the rescanned points are found to be sufficiently close to previously scanned points to be identified with them and are dropped as redundant. Point a has not been tied to any trajectory and is therefore dropped as spurious. A dummy point is created for extending trajectory B, in the hope that a satisfactory extension for it can be found in a future frame. (This is only done three consective times - then trajectory B is considered lost.) Note now that no test has been performed on whether point b might have been a valid extension of trajectory B, and whether point a might have been a valid extension of trajectory A. Assume that they would have been. Then, had an attempt been made to extend trajectory B before trajectory A, point b would have been tied to trajectory B and point a subsequently to trajectory A. Instead, extending trajectory A first tied point b to it and caused trajectory B to be broken.

The result of these problems in matching trajectories of vehicles read by a human operator was a high error rate and the need for several months work in making manual corrections of those machine errors that were detected. The implications for an automatic system are even more serious. To search for vehicle locations rapidly and with any degree of success, the automated system needs a reasonably reliable prediction of the next location of each given vehicle. It would not be unreasonable to make these regions so small that some fraction (perhaps as much as 5-10%) of the vehicles would be outside of them, because human intervention from the operator could be requested in such cases. However, the high rate of trajectory confusion experienced by the system that had available all the vehicle locations, both historical and current, makes it seem unlikely that reliable predictions of current locations could be based only on historical data. This, however, is essential for the operation of the automated system as conceived.

5. SUMMARY AND CONCLUSIONS

The results of an investigation of the techniques for automating the reduction of traffic data contained in sequences of aerial photographs of highways are reported here. It is concluded that if large amounts of traffic data are to be gathered by means of aerial photography, an automated process of reducing them to useful form is of interest both to reduce the time before the data are available and the cost associated with the manual reduction process.

The process proposed is one involving a human operator interacting with the automatic equipment. The human operator performs the more difficult tasks of pattern recognition - identifying the known ground reference points and identifying shapes as vehicles the first time they come into view. The machine performs the tasks of calculation - deriving the mapping between film coordinates and ground coordinates, record-keeping, accumulating the data on vehicle positions,and tracking vehicles through successive frames of photography. Techniques of pattern recognition are used in matching the scanned vehicle shapes in successive photographs to determine their exact position, but these are relatively simple because the most difficult task - that of identifying a shape as a vehicle - is performed by a human.

The hardware requirements of the system are a flying spot scanner, a relatively small computer with disk storage, a computer-driven display capable of displaying pictorial data from the computer, and a computer interactive graphics tablet and stylus. The computer and the associated display equipment and tablet are standard items readily available from a number of manufacturers. The specifications for the flying spot scanner are very severe, but they do not exceed those of equipment that has been built and marketed commercially.

The most important requirement for flying spot scanner is that it have the highest possible resolution. An 8000 x 8000 raster on a 70 mm film frame is required, with the highest

possible modulation transfer function. Scanning speed should be the highest achievable that is compatible with 5-bit accuracy. Raster generators external to the computer to drive both the scanner and the display, and direct memory channels between the computer and these devices for the image intensity information are desirable for speed of operation. Even though the aerial photographs used for data reduction by human operators are in color, and color information would probably be of value in an automated system as well, considerations of time required to process digitally the distinct color-separated images imply that the use of color information must be foregone. Thus, the scanner need not have a color processing capability. However, significant benefits are obtained if the film can be rotated relative to the scanning raster. The film drive for the scanner should be a precision 70 mm motorized film drive, capable of being actuated by the computer.

The major software requirements are the following:

a. Operator-computer interaction routines to identify features in the photographs. Satisfactory prototypes for these were developed at TSC in assembly language for the PDP-516 computer.

b. Programs to perform the mapping between film coordinates and ground coordinates. These are complex, but various versions exist elsewhere. A FORTRAN coded program developed under FHWA sponsorship at UCLA is directly applicable.

c. Programs to perform image correlation. These were developed at TSC in PDP-516 assembly language. Tests with these indicate that it should be possible to align vehicle images to within some 3 grid square elements, leading to positional accuracies of about $\pm$ 2 or 3 feet. This is roughly the same as the accuracy obtained by human film readers reading large numbers of vehicle positions. The speed of the search process is a critical consideration. If the region to be searched for the vehicle can be kept sufficiently small, a signifficant increase over the speed of a human operator can be achieved.

d. Programs to extrapolate vehicle paths in order to define the regions to be searched for a vehicle and to construct trajectories. The success of an automated film reading system depends on the development of improved programs to perform these functions. The routines used in conjunction with the manual film reading system are inadequate, in that they have excessive error rates even in matching vehicle trajectories when all vehicle positions have been supplied by the human operator. They project regions of possible locations for a vehicle in the next photograph that are so large that attempts to search them for the actual location using the correlation methods would be too time-consuming. A fundamental re-design of the algorithms and programs involved is required. Unless very marked improvement in their performance is achieved, the concept of automated film reading as developed here can not be usefully implemented. The development of improved path prediction and trajectory linking algorithms can proceed without reference to automatic scanning. The results would be applicable to the investigation of traffic behavior, regardless of how the data specifying individual vehicle positions at successive fixed instants of time were gathered.

Assuming an adequate path extrapolation algorithm is developed, the cost of implementing an automatic data reduction system of the type described is estimated to be somewhat in excess of $500,000. Of this amount, somewhat more than half represents the cost of the scanner and is difficult to estimate precisely, since the scanner is a one-of-a-kind device with various possible options, as spelled out above. It should be possible to procure the computer and standard peripherals for some $100,000 and two man-years of effort would be necessary to integrate the various programs into a working system. Clearly, the total investment is large and should be undertaken only if enough use for the system is foreseen to justify it. It should not be undertaken until the path-extrapolation algortihms have been developed and tested, since these are essential to successful implementation.

REFERENCES

1. Keller, M. and G. C. Tewinkel, "Space Resection in Photogrametry," Coast and Geodetic Survey Tech. Bulletin, No. 32, September 1966.

2. Fischler, M. A., "The Detection of Scene Congruence," Lockheed Missiles & Space Company Technical Report, 6-83-71-2, January 1971.