DR. ATEF A. ELASSAL*
Raytheon Company
Alexandria, Va. 22304

Data Edit Using Modular Computer Programs

The success of very large simultaneous photogrammetric solutions depends on an effective means for editing the input data.

INTRODUCTION

IT IS DIFFICULT TO CONCEIVE a successful Analytical Photogrammetric Data Processing System without an effective data editing package being associated with it. This fact becomes more evident when viewed in the light of the present days tendency towards the simultaneous processing of large photogrammetric nets.

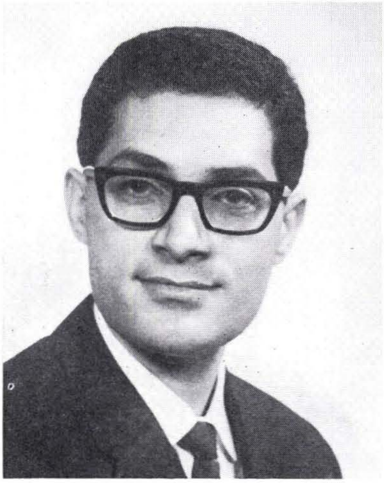
In order to arrive at a convenient form for such a data editing package, the general problem of data editing is reviewed, a logical development of some possible solutions is attempted, and the specific structure of a proposed data editing system is outlined. Finally, an illustrative example of a possible solution to one of the proposed system components is demonstrated.

REVIEW OF THE DATA EDITING PROBLEM

In the early development of most of the modern treatments of the problem of analytical photogrammetry, data editing procedures were approached as a natural extension of the least-squares treatment. The computed residuals in a problem were usually edited by comparing them to a certain multiple of the *a posteriori* statistical estimate of their standard deviations. The theoretical validity of such treatment, and the apparent ease of implementing it, produced a general concensus of its approval. The approach, as it stands, required the arrival to a least squares solution before any data editing could be performed.

It was not until the later development of large analytical data processing systems that the problems resulting from multi-source data collection systems and the associated interface problems became evident. Errors caused by these problems are usually gross in nature. Failure to identify them could result in instability of the least squares iterative procedure preventing its successful conclusion. It then follows that screening of these gross errors must necessarily precede the simultaneous treatment of the data.

Manual edit procedures used as part of quality control schemes could be injected at strategic points of the data collection and preparation system. This will undoubtedly reduce the frequency of the occurrence of blunderous errors but it will not eliminate them. It is then paramount that an automatic edit procedure should be attempted. This could only be achieved, with the re-



DR. ATEF A. ELASSAL

* Presented at the Semi-Annual Convention of the American Society of Photogrammetry, St. Louis, Mo., October 1967, under the title, "The use of modular computer programs to edit data for very large simultaneous photogrammetric solutions."

quired degree of infallibility through the use of computer controlled operations.

The question is then, "How to bring the power of the computer to tackle efficiently the problem of data editing?".

Problem definition

In order to define our objectives clearly, we will first classify accidental errors in the basic observations into two classes:

a. *Errors which exceed a certain multiple of the a priori estimate of the standard deviation of an observed quantity.* These errors are of such magnitude that they cannot be explained as a result of the stochastic sampling process. Human blunders and machine malfunctions are the largest contributors to this type of error. It is also this kind of error that causes the greatest difficulty in practice. The reason could be understood by examining the nature of the mathematical models which describe most of the photogrammetric situations. These models are usually non-linear in the unknown parameters. As least-squares adjustment could be only performed on linear systems, linear approximations of those models are used instead. The linear approximations are usually obtained through a Taylor expansion of the original mathematical model around a point in the n-space defined by the problem's parameters. The expansion is accomplished by neglecting second and higher order terms in the parameter residuals. This can be valid only if the point around which the expansion was carried is close enough to the model's solution. The failure of this condition due to the existance of blunders in the data could result in the instability of the subsequent least-squares adjustment of the derived linear model. The result of this is the failure of the solution to converge. It is quite difficult to state in general the size of the errors that will result in such instability. The geometrical situation and the degree of redundancy in the observations play a decisive role in determining the error size that can be tolerated. Fortuntaely, the majority of photogrammetric systems reflects a strong geometry by design. A fair degree of redundancy in these situations makes it possible to tolerate a rather large size of errors in the data without upsetting the stability of the least squares solution.

b. *Errors which are within the probable accuracy limits of the data collection system.* This type of error requires a rigorous statistical treatment. A test of significance which could be built around the least-squares adjustment procedure is usually used. The implementation of such tests does not represent any particular data-processing problems and, thus, will not be treated in this paper.

We will now proceed to examine methods to deal with the first type of error, described above under paragraph *a*. The data processing aspects of the problem will be treated in detail with no reference to any specific mathematical formulations that might be used to perform the edit.

Abstract: *The practical usefulness of an analytical photogrammetric data processing system is highly dependent on an effective data editing package being associated with it. The programming complexity of such a system is greatly enhanced through the use of modular computer programs. The technique makes use of a logical modulation of the photogrammetric reconstruction process to its basic components. Errors detected in the data during the execution of the different reconstruction steps are properly indexed. A logical interpretation of these indices can then be used to identify data errors and their possible causes. The proposed system structure makes it particularly useful for on-line data collection and verification operations.*

Computer Approach to the Problem

A photogrammetric situation could be reconstructed through the application of an ordered set of mathematical operations. Each one of these operations constrains the data in some way in order to satisfy a set of physically known geometrical relations. Examples of such operations in the classical triangulation case are:

- Reduction of the plate measurements to a form which reflects the principle of central projection.
- A reconstruction of the internal geometry of the photogrammetric net through the process of relative orientation.
- Scale transfer between different parts of the reconstruction.
- Formation of a photogrammetric model of the photographed object by simultaneous intersection of conjugate rays.
- Absolute orientation of the photogrammetric model in order to express the results in the desired reference system.

Failure of any of these operations or the detection of excessive residuals will point to the existence of errors in this part of the data that is being operated on. The nature of the operation being performed will also provide the clue to the possible causes of these errors. Error detection and isolation procedures could be built in these operations. Each operation must properly index that part of the data found in error so that it will be identifiable to any subsequent operation. Because certain types of errors are unidentifiable (masked) to some operations, the sequence of application is important and could be used to the advantage of increasing the effectiveness of the data edit process.

The data edit system must provide for the applications of any set of operations in any desired sequence. This degree of flexibility is imposed on the system for a number of reasons:

- ★ Problems to be handled will reflect a wide variety of geometrical configurations.
- ★ In the case of hybrid systems, provisions must be made to add or delete photographic or control information.
- ★ In case of on-line data-collection and verification, the system must be capable of handling unscheduled accumulation of any type of data.

This required degree of flexibility could be achieved by employing modular programming technique.

Modular Programming

It is possible to break down any data processing system into a number of logically independent operations. Each one of these operations transforms the problem data to a higher level of processing until the complete processing is achieved. A program is thus a preconceived assembly of a number of such operations. To facilitate the programming and assembly of these operations, they are usually programmed in a form of modules. From the programming point of view, the modules must have identical form. This means that the sequence of operations which effect a transfer of control to and from a module must be identical.

Responsibility of controlling the assembly and execution of these modules resides with a program which is usually termed the Supervisor. The Supervisor could assemble and execute a program phase consisting of one or more modules which may be arranged in any desirable sequence.

The logic involved in controlling the relay of the required data to an arbitrary arrangement of the modules could be extremely complex. A great deal of this complexity can be resolved by standardizing the data format and by only allowing a program module to change the data contents leaving its standard format intact. However, each module should be responsible to check the data in order to find out whether or not it meets the module's minimum requirements. The concept of data manipulation and formating is termed Data Management.

Data Management

A large percentage of programming time and effort are spent in setting up the mechanics of data storage and retrieval. This is particularly true in photogrammetric applications where both the amount and diversity of the data handled are usually extensive. A great deal of the efficiency of any data processing system depends on how data is manipulated between the different program modules. One convenient way of keeping track of data storage and retrieval is through the concept of data sets. A data set is defined as a named collection of data. Plate measurements are an example of such a data set. The data set in this case could be composed of image identifications, plate coordinates, covariance matrices of plate coordinates and a set of indices that indicates the data status. The plate identification could serve as the data set name. A number of data sets could be filed in a Catalog. Each entry in the Catalog contains a data set name and an indication of its location. The location indicator contains both the storage device and where the data set resides on this device. The collection of data sets and their catalogs are defined as a data Volume. In photogrammetric applications, separate data volumes could be used for plate measurements, camera station parameters, and ground control. It is to be mentioned that the cross reference between the different volumes of data is usually carried through camera-station and point identifications. It should be also noticed that the status of each piece of data is provided by an indexing system which is part of the corresponding data volume.

The index of an element of data must be able to indicate the set of conditions that the element could assume. For instance, the index of an image-point coordinates may indicate whether the point is active or has been dropped. The same index can also reveal the program module which rejected that point. In the case of the camera station parameters, the index system could also reflect the type of coordinate system in which they are expressed A translation of the index system at any time will provide a complete picture of the data status. Storage could be employed most efficiently if special number systems were used for each type of indices. The basis for these numbering systems will be equal to the various conditions the corresponding piece of data may assume. Thus, a binary number system will be used to reflect a binary condition, and so on.

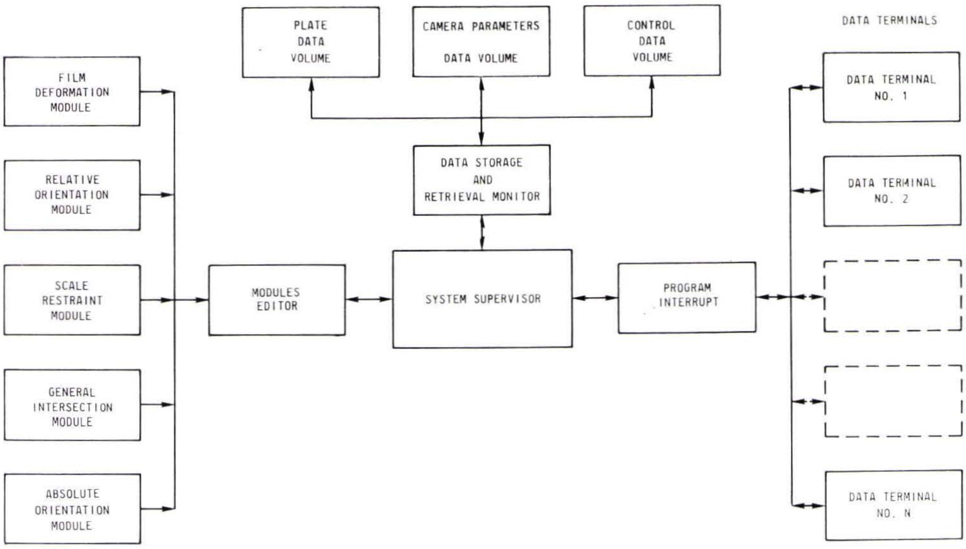


FIG. 1. Proposed organization chart for data varification and filing operation system.

The concept of data management as given above is intended to give a general idea of the problem extent. The details are far more complicated and allow the system designer ample room for a great deal of creativity. The point to be emphasized here is that, once a data management system is established, the program modules using it must not be allowed to change its form. This arrangement will make possible an arbitrary sequence of application of the program modules.

The system could be designed to operate in one of two modes:

Off-Line Operations. This mode of operation may be used if data verification is to be performed after the process of data collection is completed. This will be necessary if a computer is not accessible to handle the data during its accumulation.

The module-operating sequence can be generated by the supervisor program or by the user as part of his input. Program-generated sequence will be possible only in case of regular photographic coverage where logical procedures could be employed to generate the sequence.

The results of this operating procedure will be a complete diagnosis of the verified data pointing the errors and their possible causes.

On-Line Operations. This type of operating procedure will fit the needs of fairly large organization where the corresponding large volume of data handled requires efficient use of its computing facilities. The system will be operating on a time sharing basis with different terminals feeding the computer, possibly in a simultaneous fashion. Data-file updating and error detection will then be conveniently kept in step with the data acquisition process. Error rectification could be achieved as the data is accumulated.

Figure 1 shows the proposed system's organizational chart.

ILLUSTRATIVE EXAMPLE

The principles stated here were attempted on an experimental scale in a program for cantilever extension developed at Autometric/Raytheon. The program revolves around two program modules; relative orientation and scale restraint. A wide variety of operating options were built in the program in order to give the user a greater control over the data flow. These options in terms of data editing are:

AUTOMATIC MODE

In this mode, the program performs data editing automatically, guided with control parameters given to it by the user. These parameters, augmented with others which are internally generated, provide the norms for the data rejection process. For instance, in the relative orientation process, the user provides the program with a rejection criteria for plate coordinate residuals. The user also supplies the program with the minimum degrees of freedom that have to be retained in the system. The plate coordinate rejection criterion ϵ_1 will reflect the user's estimate of the maximum acceptable residuals based on his evaluation of the material, equipment and procedure used in collecting the data. Another plate coordinate rejection criteria ϵ_2 is a certain multiple of the *a posteriori* estimate of the plate coordinate standard deviation.

TABLE 1. IMAGE COORDINATES OF POINTS SHOWN IN FIGURE 2

Point ID	Left		Right	
	X (mm.)	*Y (mm.)*	*X (mm.)*	*Y (mm.)*
UA	0.000	110.000*	−105.000	100.000
UB	50.000	100.000	−55.000	100.000
UC	100.000	100.000	−5.000	100.000
MA	0.000	0.000	−105.000	0.000
MB	50.000	0.000	−55.000	0.000
MC	100.000	0.000	−5.000	0.000
LA	0.000	−100.000	−105.000	−100.000
LB	50.000	−100.000	−55.000	−100.000
LC	100.000	−100.000	−5.000	−100.000

* Correct Value = 100.000
Principal Distance = 150.000 mm.

This value is computed by the program for each attempted relative orientation solution. Plate residuals are first checked against ϵ_1. If any rejections are detected under this test, it will be limited to those points with the highest residuals provided that the requirement for minimum degrees of freedom is satisfied. The test is always performed on all points in order to provide means for re-including points that were previously rejected but later were found to be acceptable. The data edit process continues until a stable condition is reached where no points are either rejected or included. When this happens, the program proceeds to perform the same operations again using ϵ_2 instead of ϵ_1. The program will shift to manual mode if the conditions of the edit were impossible to fulfill or where the number of editing trials exceeds a preassigned number.

MANUAL MODE

In this mode the program relinquishes control to the user at certain decision points. The user can then relay instructions about what course of action that should be taken through the combined use of console typewriter and sense switches. The user can also change rejection criterion and return to the automatic mode if he wishes.

The manual mode was found to be necessary in order to provide a means to deal with marginal cases which require intelligent decisions which cannot be built in the program logic without undue complications. These cases arise primarily where weak geometry or low degree of overdetermination exists.

We will now proceed to illustrate the mechanics of the data verification as exercised by the program. A simple example of a single stereomodel is used. (See Table 1, Figure 2). The model simulates two vertical photographs with a total of nine points in their common overlap. An error of 10.000 mm. is introduced in the y-coordinate of one point. The program was supplied with the following parameters:

Maximum allowable plate residuals 0.050 mm.
Minimum degrees of freedom 2.
Maximum allowable number of trials 5.

Three trials were required to isolate the error. Table 2 and Figures 3a through 3c illustrate a computer printout of the data status after each trial. In Figure 3a, the plate

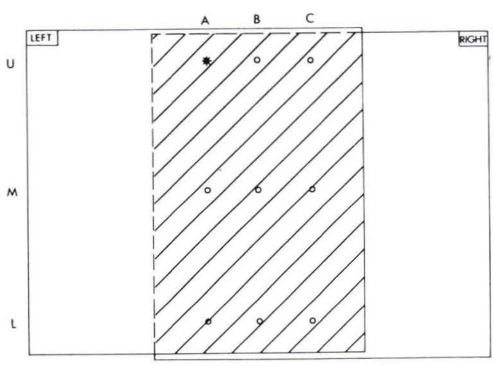


FIG. 2. Single stereoscopic model and arrangement of points associated with Table 1.

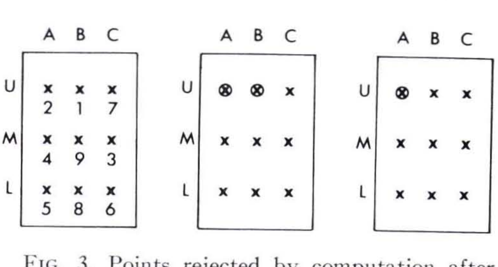


FIG. 3. Points rejected by computation after first, second and third trials as shown in Table 2. (Figures 3a, 3b, and 3c are numbered left to right.)

TABLE 2. PRINTOUT OF PLATE RESIDUALS FOR THREE TRIALS ASSOCIATED WITH FIGURES 3a, 3b, AND 3c

	Plate Residuals of Model Left				*-Right Type*	*Ser No*
ID	*DX1*	*DY1*	*DX2*	*DY2*		
UA	0.070	1.105	0.025	−1.192		1
UB	−0.101	−1.598	−0.031	1.680		2
UC	0.025	0.394	0.008	−0.404		3
MA	−0.037	−0.014	0.014	0.783		4
MB	0.000	0.007	−0.000	−0.007		5
MC	0.041	0.827	−0.013	−0.820		6
LA	0.016	0.436	−0.024	−0.439		7
LB	−0.001	−0.016	0.001	0.016		8
LC	−0.016	−0.421	0.022	0.404		9
Weighted Sum of Squares=0.11617785E 02 Degrees of Freedom=4 Unit Standard Deviation=1.704						
UA	0.000	5.000	−0.000	−5.000	OUT	1
UB	0.000	0.000	−0.000	−0.000	OUT	2
UC	0.000	0.000	−0.000	−0.000		3
MA	0.000	0.000	−0.000	−0.000		4
MB	0.000	0.000	−0.000	−0.000		5
MC	0.000	0.000	−0.000	−0.000		6
LA	0.000	0.000	−0.000	−0.000		7
LB	0.000	0.000	−0.000	−0.000		8
LC	0.000	0.000	−0.000	−0.000		9
Weighted Sum of Squares=0.20761789E-16 Degrees of Freedom=2 Unit Standard Deviation=0.000						
UA	0.000	5.000	−0.000	−5.000	OUT	1
UB	0.000	0.000	−0.000	−0.000		2
UC	0.000	0.000	−0.000	−0.000		3
MA	0.000	0.000	−0.000	−0.000		4
MB	0.000	0.000	−0.000	−0.000		5
MC	0.000	0.000	−0.000	−0.000		6
LA	0.000	0.000	−0.000	−0.000		7
LB	0.000	0.000	−0.000	−0.000		8
LC	0.000	0.000	−0.000	−0.000		9
Weighted Sum of Squares=0.15813595-15 Degrees of Freedom=3 Unit Standard Deviation=0.000						

residuals did not meet the rejection criteria of .050 mm. The program then rejected the two points *UA* and *UB* which exhibit the largest residuals and at the same time retaining the required 2 degrees of freedom. The program after the second trial (See Figure 3b) accepts the previously rejected point *UB*. Figure 3c presents the last trial in which the residuals are found to meet all the data editing criterion.

CONCLUSIONS

The modulation of the photogrammetric reconstruction process could be effectively employed to produce a highly flexible data editing system. The structure of such a system makes it particularly useful for on-line data collection and verification operations. The extent of such system could be altered with relative ease to make it fit a wide range of organizational needs.