SHU - AN INTERACTIVE GRAPHICS PROGRAM FOR THE STORAGE, RETRIEVAL AND ANALYSIS OF ARTEFACT SHAPES

P.L. Main

British Museum Research Laboratory, Great Russell Street,
London WC1B 3DG

Brief description of SHU

SHU is an interactive graphics program, written in Algol 60, and currently implemented at the British Museum Research Laboratory on a Hewlett-Packard mini-computer. The program occupies at present 52K bytes in memory, but part of this area is overlayed. The total code of the program (root + 5 segments) is about 120K bytes. It is running under the following configuration:

Hewlett-Packard 1000 minicomputer with RTE-IVB operating system,
120 Mbyte hard disc, Line/Matrix printer, 2 magnetic tape drives,
4-colour flat-bed graph plotter and VDU terminal.

Of these the printer and magnetic tape drives are not essential to the program's operation. The system would be considerably enhanced by an interactive graphics VDU, but this is not available in the British Museum installation at present.

The program is designed to satisfy the needs of two types of user. At one level, it can be used for storing the outline shapes of artefacts in files, reproducing them on a graph plotter, and adding lines, text, etc. to form a complete illustration. This illustration can be stored permanently as a file on disc. Only a familiarity with appropriate program commands is necessary for these sorts of operation - no knowledge of the internal operation of the program is required. All the illustrations in this paper are generated by SHU command sequences. At a slightly deeper level, SHU can be used for the comparison of shapes using a number of different similarity measures. Similarity matrices between groups of artefacts can be calculated and simple retrieval operations performed. For these purposes of course, a clearer knowledge of how outline shape is stored and compared by the program is required.

Program Command Format

SHU commands can come from either the VDU terminal on which the program is running, or a command sequence can be stored as a text file on disc. We will refer to such a file as a command file. The format of the command is best illustrated by an example. The command sequence in Table 1 was used to plot the illustration in Fig. 1. The precise function of each command is not considered here - but each begins with a four-letter mnemonic describing the effect of the command, followed by a series of parameters separated by commas. If those parameters appearing within the boxes in Table 1 were removed, this would represent the command sequence as originally typed on the user's terminal. When typing commands from a terminal, it is possible to specify that the commands typed are to be simultaneously written to a disc file (the echo file) as they are entered. This has two important benefits. Firstly, the echo file remembers the commands which were given to create an illustration, and so if this file is given to SHU on a subsequent occasion as a command file, a duplicate illustration will be produced. To appreciate the second benefit, consider the first PLAR (PLot ARtefact) command in Table 1. This would be typed in as:

PLAR,2,,,4

which, roughly speaking, instructs SHU to draw the second artefact from the input file on the plotter. Evidently, the program needs to know where the artefact is to be plotted. At this point the program interacts with the user by instructing him to move the plotter pen manually to the point where the artefact is to appear, and waits until a button on the plotter is pressed. The coordinates of this point are then appended to the PLAR command before it is written to the echo file, and all the parameters appearing within the boxes in Table 1 represent positions on the plotter table which have been appended in this way. This echo file is used as a command file, subsequently these parameters are used to determine the position of the various elements of the illustration, and a reproduction of the original is produced without any intervention by the user.

Table 1: The SHU command file used to plot Fig. 1.

```
INFL, RIMO2
OFST,-0.8,-2.0
RALS,ILLALS
CSIZ,0.12
SCAR,2.5
PLAR,2,,,4,  2.259, 6.446
PLAR,14,,,4, 3.032, 6.446
PLAR,22,,,4, 3.732, 6.446
PLAR,30,,,4, 4.465, 6.446
PLAR,11,,,4, 1.813, 4.970
PLAR,12,,,4, 2.528, 4.970
PLAR,13,,,4, 3.230, 4.970
PLAR,15,,,4, 3.958, 4.970
PLAR,26,,,4, 4.808, 4.970
PLAR,4,,,4,  1.810, 3.598
PLAR,5,,,4,  2.536, 3.598
PLAR,23,,,4, 3.325, 3.598
PLAR,24,,,4, 4.073, 3.598
PLAR,32,,,4, 4.838, 3.598
PLLI,, 1.503,5.711,5.041,5.711
pltx,,,, 6.000, 6.530
Group 3 (I.of Wight)
plli,, 1.531, 4.423, 5.061, 4.423
pltx,,,, 5.977, 5.022
Group 4 (Wilts.)
plli,, 1.536, 3.064, 5.116, 3.064
pltx,,,, 5.975, 3.729
Group 5 (Berks.)
OFST,-0.8,-1.3
CSIZ,0.12,0.55
PLTX,2,,, 1.537, 1.671
Fig. 2. The RIMS dataset arranged
according to the
pltx,2,,, 2.600, 1.460
excavation reports in which they
appeared.
/E
```

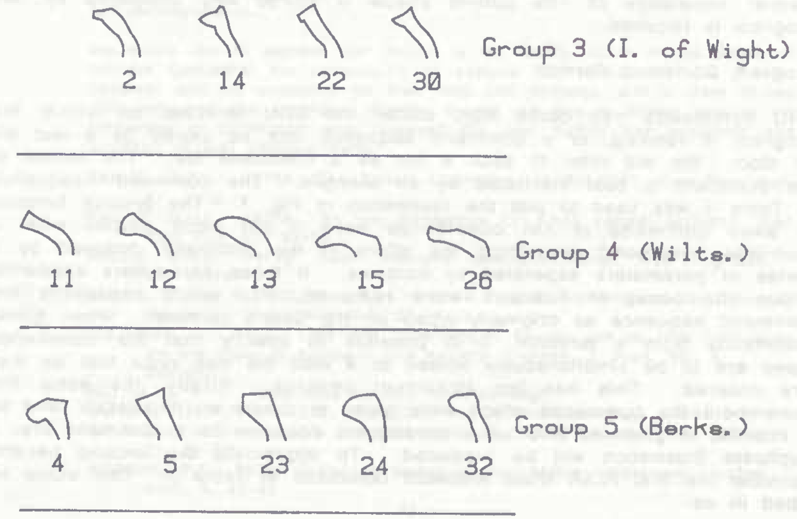


Figure 1: The RIMS dataset arranged according to the excavation reports in which they appeared.

In short, therefore, the use of echo files provides a powerful method for interactively building up illustrations from individual elements and storing the result as a permanent file. Quite complex illustrations can be pieced together in this way. Furthermore, since echo files are text files, layout alterations can easily be made using a text editor.

Recording artefact shapes

SHU is suitable only for the recording of two-dimensional outline shapes. Typically this will come from a drawing or a photograph. Three-dimensional information can only be accommodated in the form of a series of two-dimensional sections through an artefact in different planes. It assumes that shape is recorded as a sequence of digitised points lying sequentially along the outline, starting at some arbitrarily defined 'reference point'. A possible scheme for digitising the minimum necessary shape information from a medieval cooking-pot outline is illustrated in Fig. 2. Here the digitising sequence is from the reference point A, via points F, Q, R and ending at G. The methods provided for expanding this information into a presentable drawing of a cooking pot are described in the next section. SHU does not itself provide digitising facilities, but reads the coordinates from a text file on disc. The module which reads this information has been carefully isolated and can easily be replaced to suit the output format from any digitising equipment. Table 2 shows the coordinate sequence corresponding to Fig. 2, and is in the format produced by a d-Mac digitising table. The information within boxes has been added manually during the digitising process, and its significance is discussed below.

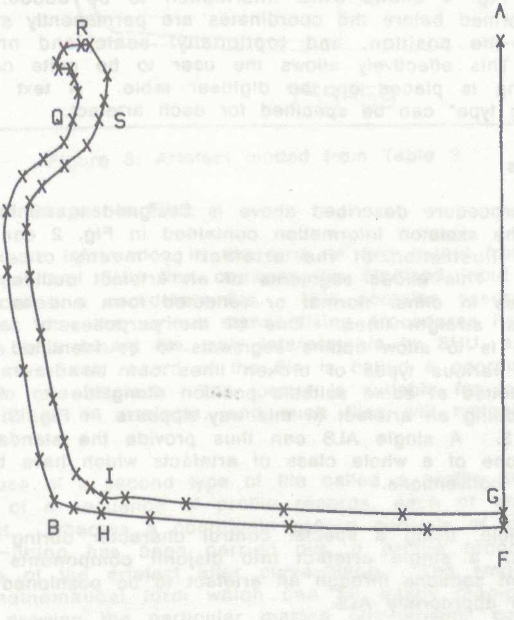


Figure 2: An example of a scheme for digitising a mediaeval pot.

Table 2: Digitised coordinate sequence corresponding to Fig. 2.

```
'755-59'            X4901 Y0492     RX4944 Y0625      X4927 Y0417
4.0    4            X4901 Y0528     ↑X4949 Y0625      X4933 Y0392
 AX5185 Y0626       X4910 Y0548     ↑X4959 Y0607      X4940 Y0373
F↑X5185 Y0340       X4922 Y0558      X4957 Y0592      X4948 Y0365
  X5141 Y0339       X4934 Y0568     SX4954 Y0581     IX4956 Y0359
  X5062 Y0341      QX4939 Y0581      X4947 Y0566      X4968 Y0359
  X4982 Y0347       X4942 Y0597      X4934 Y0554      X5003 Y0355
 HX4954 Y0350       X4940 Y0605      X4921 Y0542      X5059 Y0351
  X4938 Y0353      ↑X4938 Y0611      X4914 Y0533      X5127 Y0348
B↑X4927 Y0357      ↑X4931 Y0611      X4912 Y0517     GX5185 Y0348
  X4919 Y0382       X4930 Y0615      X4914 Y0489    -1
  X4910 Y0434       X4934 Y0623      X4920 Y0455
```

When SHU plots artefacts, or makes comparisons between them, it performs smooth curve-fitting between the sequence of digitised points. This means that relatively few points need be digitised to produce a visually acceptable plotted outline (see Fig. 2). It does, however, mean that a facility must exist for suppressing the curve-fitting at specified points, i.e. where there is a corner on the original drawing. Corner points are indicated during digitising by preceding the coordinate pair with a '^' symbol as can be seen in Table 2. Furthermore, any digitised point can be given a one-character label, so that the point can be uniquely identified later. In this case the coordinate pair is simply preceded by that character. Table 2 contains labelled points A, F, Q, R, S and G. The other manually entered information in Fig. 4 allows extra information to be added, or certain operations performed before the coordinates are permanently stored by SHU. For example the position, and (optionally) scale and orientation are standardised. This effectively allows the user to be quite careless about how the drawing is placed on the digitiser table. A text label and an integer "drawing type" can be specified for each artefact.

Plotting artefacts

The digitising procedure described above is designed to minimise the effort required, and the skeleton information contained in Fig. 2 can be built into a presentable illustration of the artefact by means of artefact layout schemes (ALS). This allows segments of an artefact outline to be plotted at various scales in either normal or reflected form and to be connected up by additional straight lines. One of the purposes of labelling points during digitising is to allow outline segments to be identified within ALSs. Either solid or various types of broken lines can be drawn; the artefact label can be plotted at some suitable position alongside it. An example of the result of plotting an artefact in this way appears in Fig. 3, with the ALS used in Table 3. A single ALS can thus provide the standard format for displaying any one of a whole class of artefacts which have been recorded using the same conventions.

It is also possible, using a special control character during the digitising process, to split a single artefact into disjoint components allowing, for example, different sections through an artefact to be positioned together and plotted using an appropriate ALS.

Table 3: Artefact Layout Scheme used to plot Fig. 3.

```
LI,A,F,1         (line from A to F)
CU,F,H,2         (pot outline from F to H with broken line)
CH,H,I,1         (outline from H to I with solid line)
CU,I,G,2         (I to G with broken line)
LI,I,G,0         (move pen from G to F)
NA,0,-0.8,5      (plot artefact label)
CU,F,H,2,X       (plot reflection of F to G with broken line)
CU,H,R,1,X       (reflection of H to R with solid line)
LI,A,R,1         (line from A to R)
LI,A,R,1         (and again)
LI,-2,-1.5,0     (move pen off to left)
CU,S,Q,1,,2.0    (plot rim at twice scale)
/E
```

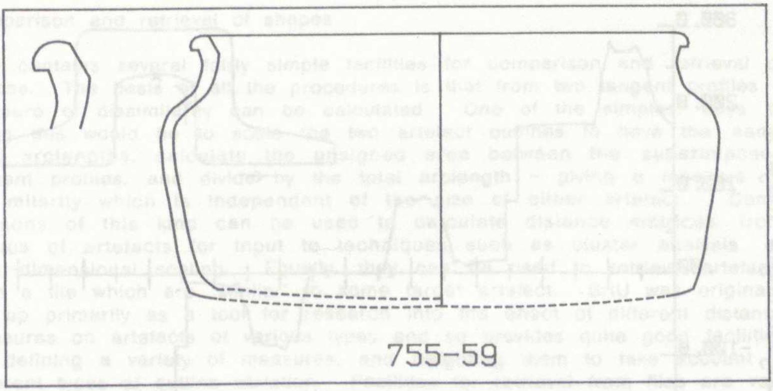


Figure 3: Artefact plotted from Table 3.

Storing Artefact Shapes in Files

Storing coordinate information in the form of text output from a digitiser is rather inefficient and SHU first converts the digitised input file to another type of file called a coordinate file. This contains essentially the same information, but after the various standardising procedures have been carried out. The file is a binary file, only interpretable by SHU, and is in a very compact format. Each record of the file is called a coordinate record and corresponds to one artefact. The format is suitable for long-term storage of large quantities of artefacts, and such files will typically be held on magnetic tape.

SHU makes use of a second type of file called a profile file. Once more, this consists of a sequence of profile records, each of which corresponds to an artefact. Whereas a coordinate record consists of shape information before curve-fitting has been carried out, a profile record contains the outline profile of the artefact after curve-fitting. The smoothed profile is held in a mathematical form which can be easily manipulated, for the purposes of drawing the particular making comparisons between shapes. The form used is a tangent profile, i.e. a mathematical function of tangent angle against arclength. Arclength is measured from the 'reference point', that is to say the first digitised point on the artefact, and measurement of tangent angle proceeds along the smoothed outline in the original direction of digitising. An example of a tangent profile and the artefact to which it corresponds appears in Fig. 4. Although a profile record is in a very convenient form for mathematical operations, it occupies approximately four times the space of the coordinate record from which it is generated. Profile files are thus normally created as temporary disc files for the purpose of a particular analysis and later discarded. The benefits of using two types of file for storing and analysing shape has already been discussed in Main (1979).

A diagrammatic representation of the successive transformations undergone by the original digitised input record is given in Table 4. To the user who wishes only to record and plot his artefacts, the process of transforming coordinate records to profile records and integrating them to provide the

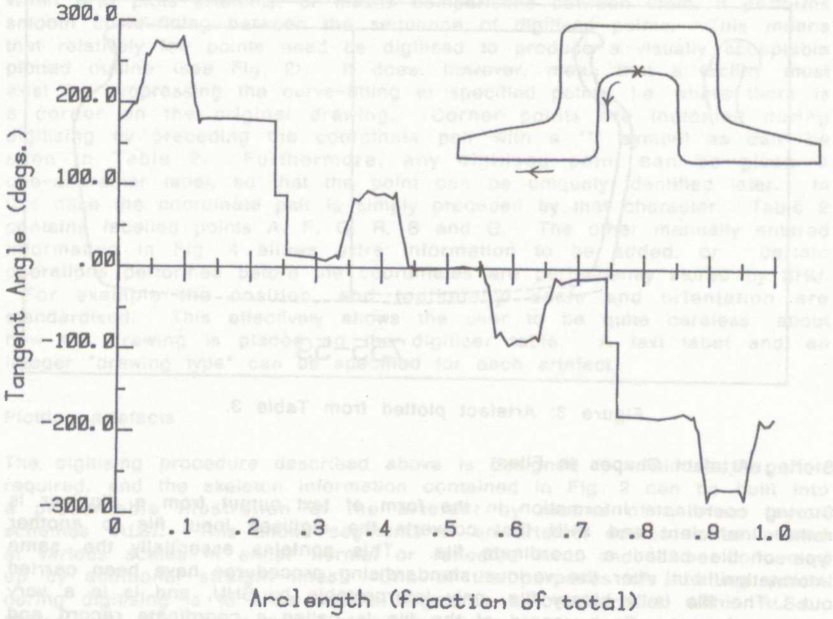
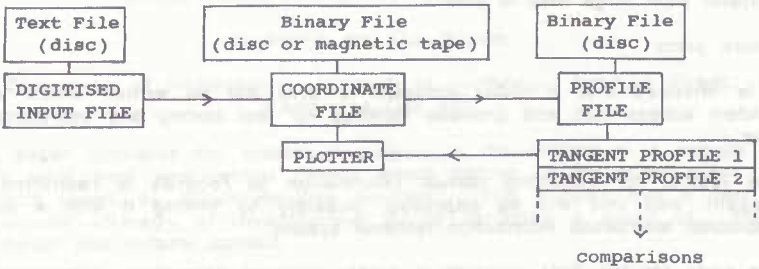


Figure 4: Section through a 19th century railway-line and its tangent profile. The cross shows the origin of the arclength and the arrow the direction of increasing arclength.

2-dimensional representation is quite invisible. To him it appears that he is plotting an artefact directly from a coordinate file (indicated by the dotted arrow in the table). For the purpose of analysing an entire dataset, however, it would be most efficient to create a profile file containing the required artefacts. Of course to plot or compare artefacts, the tangent profiles concerned must be read into memory - and the overall efficiency of the system is greatly affected by the number that can be held there at one time. The limited addressing capability of the Hewlett-Packard 1000 system means that this is quite severely restricted, but on a mainframe far more could be accommodated.

Table 4: Storage of shape information within SHU.



Comparison and retrieval of shapes

SHU contains several fairly simple facilities for comparison and retrieval of shapes. The basis of all the procedures is that from two tangent profiles a measure of dissimilarity can be calculated. One of the simplest ways of doing this would be to scale the two artefact outlines to have the same total arclengths, calculate the unsigned area between the superimposed tangent profiles, and divide by the total arclength - giving a measure of dissimilarity which is independent of the size of either artefact. Comparisons of this kind can be used to calculate distance matrices from groups of artefacts for input to techniques such as cluster analysis or multi-dimensional scaling. Equally, they can be used to retrieve artefacts from a file which are 'similar' to some target artefact. SHU was originally set up primarily as a tool for research into the effect of different distance measures on artefacts of various types and so provides quite good facilities for defining a variety of measures, and weighting them to take account of different types of outline variation. Facilities for retrieval from files are very basic, however, and plans exist to implement some more efficient sorting and searching algorithms using these distance measures.

It should be stressed that distance calculations of the type outlined above do not rely on taking measurements or ratios of measurements between 'features' of the artefacts and so comparisons can be made even on datasets of artefacts some of which contain missing features, or even on artefacts presenting no consistent features at all, such as pot-rims.

Limitations and problems

The method of comparing shapes used by SHU has a fundamental limitation imposed by the method of representation. This means in effect that sensible comparisons cannot be made between broken artefacts unless the original outlines can be reconstructed.

SHU uses a particular class of methods of comparing shapes - i.e. some form of integrated difference between tangent profiles. On occasion this can give different results from those based on visual perception (as can any objective method). In general, however, the distance measures provided give good results over a wide range of artefact types.

At present SHU uses only sequentially accessed files, and this means that retrieval from large files is slow.

Future plans

It is intended that a sister program to SHU will be written which uses random access files and provides facilities for fast sorting and searching of files.

The facility for attaching textual information to records is restricted at present, and this will be improved, possibly by linking it with a more traditional text-based information retrieval system.

The code for the SHU program is highly machine dependent, particularly in the input/output routines. A high priority is to rewrite the program in a reasonably portable form, probably using PASCAL.

MAIN, P.L. 1979 Data-banks of archaeological shapes. Comp. Appl. Arch. 7, 5-12.