

Combining quantitative analysis with an elliptic Fourier descriptor: A study of pottery from the Gansu-Zhanqi site based on 3D scanning and computer technology

Jin Wang ^{a,*}, Wei Qian^b, Guoke Chen^c

^a Institute for history of science and technology, Shanxi university, Taiyuan, 030006, P R China.

^b Institute for Cultural Heritage and History of Science and Technology, University of Science and Technology Beijing, Beijing 100083, P R China.

^c Gansu Provincial Institute of Archaeology, Lanzhou, 730001, P R China.

* Corresponding author. Email: wjustb@163.com

Abstract: Pottery is an important material in archaeological studies, and the accurate classification of pottery shapes largely depends on the experience and knowledge of archaeologists. In this thesis, pottery taken from the Gansu-Zhanqi site is used for sampling. Three-dimensional models of the pottery were obtained by three-dimensional scanning, and a computer-assisted pottery typology was studied through quantitative analysis and elliptic Fourier analysis. This method can enhance and supplement the traditional methods of classifying pottery in archaeology, thereby enriching the parameters and breadth of pottery analysis. This method represents a new means for exploring and experimenting with objective classification and provides a new tool for traditional archaeological analysis methods.

Keywords: Pottery, Gansu-Zhanqi Site, Quantitative analysis, Elliptic Fourier descriptor, Computer programming, 3D scanning

Introduction

Pottery is a common and important ancient material in the study of archaeology. The classification pattern of each object is an important factor used to determine the type and age of an archaeological culture. Over the past few decades, some archaeologists have utilized mathematical methods such as seriation [1], principal component analysis [2], cluster analysis [3], multivariate statistics [4], correspondence analysis [5], and principal coordinate analysis [6]. Chinese archaeologists have used cluster analysis, which is a multivariate statistical method, to classify pottery ware [7].

Traditionally, these mathematical methods have been used alone, and the data

have been measured manually. In this paper, we used a computer program to extract data from three-dimensional pottery models. The data were used for quantitative analysis and elliptical Fourier analysis. Elliptic Fourier descriptors (EFDs), originally proposed by Kuhl and Giardina [8], can delineate any type of shape with a closed two-dimensional contour. Over the past several years, researchers have applied elliptical Fourier analysis to various fields, such as agriculture [9], the human body [10, 11], and ecology [12], but it is rarely used in archaeological research. We combined the results of quantitative analysis and elliptical Fourier analysis to study pottery typology. The final analysis provides more parameters, which make the final results more accurate.

In this article, we use unearthed pottery from the Zhanqi Burial site as examples. The Zhanqi site is a part of the Siwa cultural relics (from 14 BC to 11 BC) along the Taohe River in Min County, Gansu Province. Sixty-six tombs were found at the site, as well as 20 sites that included houses, cooking pits, ash pits and sacrificial remains. The unearthed objects include pottery, bronzeware, stone tools, bone artifacts and ornaments. The layers and times of the tombs are clear, and unearthed artifacts are abundant. The Zhanqi site is of great significance for the further study of the Siwa culture [13]. Considerable amounts of pottery were unearthed at the Zhanqi site, including jars, lis, basins, dous, and bottles. Most pottery consists of jars. According to the shape of the mouth, jars can be divided into flat and saddle jars [14]. In this article, 62 saddle jar examples, which are representative pottery from the Zhanqi site, were scanned.

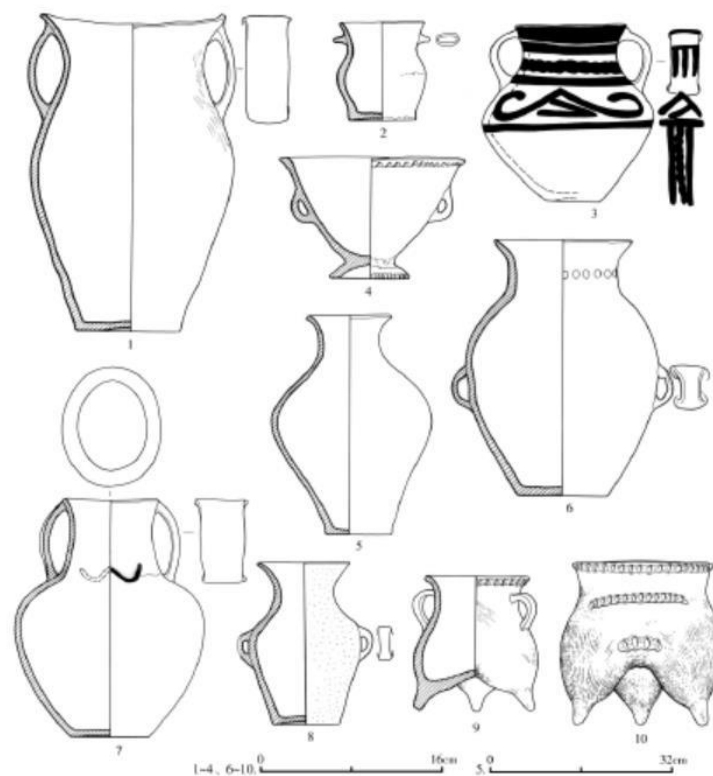


Fig. 1. The unearthed pottery ^[13]

1. Saddle jar 2. Flat jar 3. Coloured pottery 4. Dou 5. Bottle
6. Bottle 7. Flat jar 8. Bottle 9. Li 10. Li

First, it is important to note that the pottery model is created using 3D scanning. The pottery model can be used to create a 3D virtual display, print a 3D model, perform a virtual restoration of cultural relics, and record 3D data. A 3D pottery model is permanent and can be used multiple times as a basic model for further research. This article is only one part of this research. Second, the quantitative analysis of pottery has been applied to many archaeological studies. Compared with traditional archaeological methods, cluster analysis results were basically identical to the results provided by the original reports. The results of this study have been confirmed [15]. Curves can be analyzed in a variety of ways. The elliptic Fourier descriptor is a highly suitable method for this paper. The pottery of the Gansu-Zhanqi site has not been analyzed through traditional archaeological methods. The results of this pottery typology are the first to be published for this site. Therefore, the results of our study in this article provide a reference for traditional archaeology, although it cannot furnish conclusions for archaeological research.

Unifying the Data Extracted from the Pottery Model

The 3D models were acquired using a Creaform Go!SCAN 20 scanner [16]. The accuracy is 0.1 mm, and the resolution is 0.2 mm. Compared with a 3D laser scanner, this scanner uses a white light scanner with an LED light source and a fast scanning speed. This scanner is more suitable for small cultural relics. VXelements [17], Autodesk Meshmixer [18], and 3D Builder [19] are used for the postprocessing stage of the 3D scanning process. It is important to note that since the research on pottery shapes requires less accuracy than a virtual display, the model analysis is not affected by the software processes in this paper.

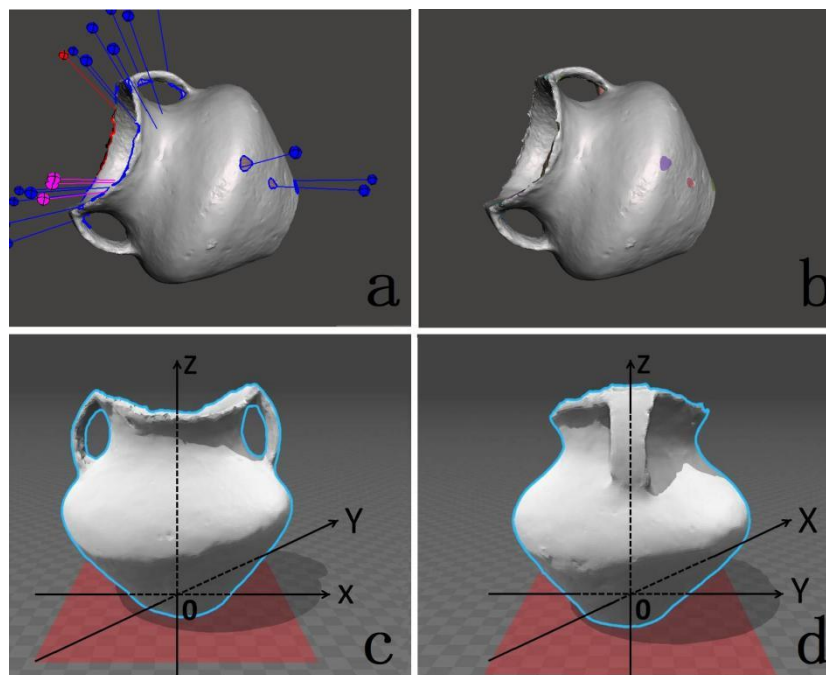


Fig. 2. Postprocessing the scanning data.

- a. Deficiencies are detected, such as holes and cracks, using the Autodesk Meshmixer
- b. The model is complete after being restored in Autodesk Meshmixer
- c. The dimensional data are based on the XZ plane in the 3D Builder.
- d. The curvature data are based on the YZ plane in the 3D Builder.

The data are extracted from the three-dimensional pottery model and include two kinds of data. The dimensional data are used for quantitative analysis, and the curvature data are used for elliptic Fourier analysis. Over the past few years, scientists have introduced several data extraction methods for three-dimensional pottery models, including Spelitz [20], Rasheed [21], Măruțoiu [22], and Angelo [23]. In this paper, the data are extracted using a computer program in C language. It should be noted that the jars at the Zhanqi site have two ears. The binaural jar is a representative characteristic of pottery from northwest China. Considering the methods and functions of this article, the extraction and analysis of the dimensional data are based on the XZ plane in the model, which shows the different calibers and heights of the ears. The extraction and analysis of the curvature data are based on the YZ plane in the model, which shows differences in the overall shapes.

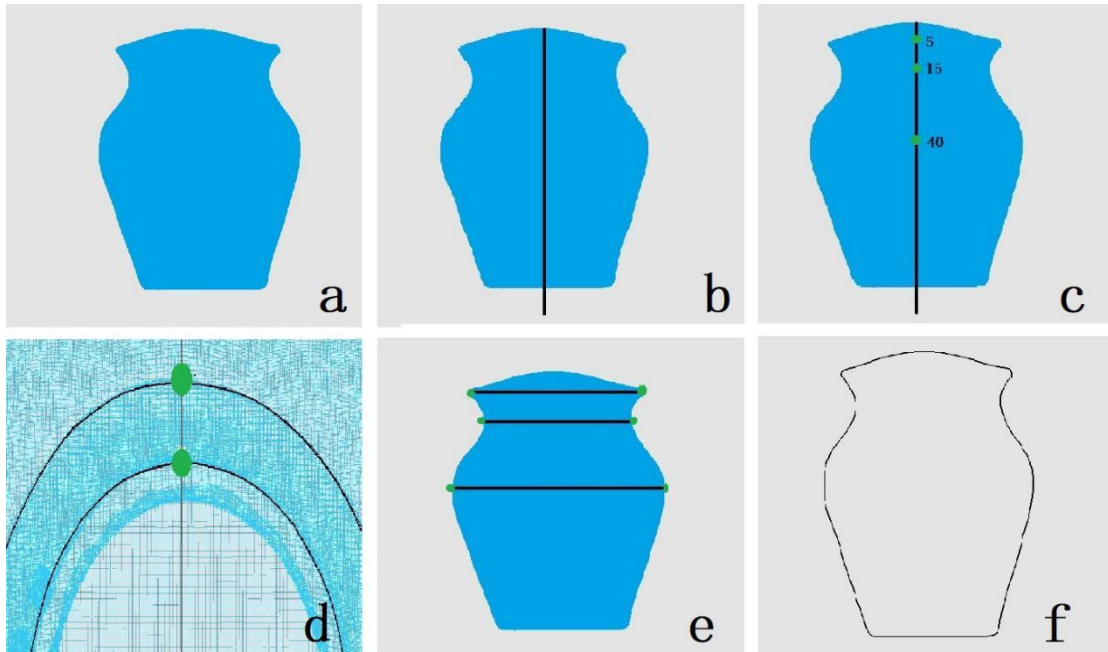


Fig. 3. Data acquisition process in the C programming language

- a. The point cloud data are read.
- b. The contour is generated.
- c. The contour points are generated according to the calculation.
- d. The contour line is enlarged.
- e. The plane is formed according to the contour points.
- f. The contour lines of the pottery model are output.

Methods

A. Quantitative analysis

Step 1. The first step is to obtain the original data. According to the three-dimensional structure of a pottery model, the areas from which data are extracted differ. To a certain extent, these areas represent the characteristics of the pottery, which are the data used for the quantitative analysis. The measurement areas are shown in **Fig. 2**.

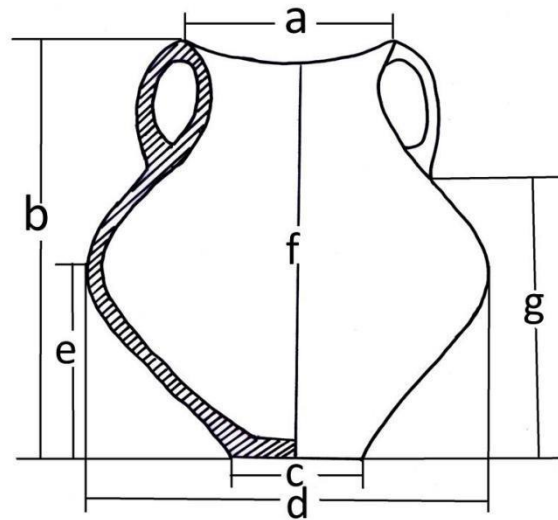


Fig. 4. Measurement areas for a jar

- a. Caliber b. Height (the highest) c. Bottom diameter
d. Abdominal diameter e. Height of the abdominal diameter
f. Height (the shortest) g. Height of the ear bottom

Step 2. The second step involves using the original data and selecting a different proportion to represent a feature of the pottery ware.

Sixty-two jars were analyzed by examining the overall structure of the pottery model, analyzing the characteristic lengths, and choosing meaningful ratios for the following:

- Abdominal diameter/height (the highest), which represents the overall shape of the object;
- Height of the abdominal diameter/height (the highest), which represents the position of the abdominal diameter;
- Height of the ear (b-g)/Height (the highest), which is the ratio of the ear height to the overall shape; and
- Height (the shortest)/Height (the highest), which is the degree of the saddle mouth.

Step 3. The third step is to conduct quantitative analysis using the ratios obtained in the second step. In this paper, we use the multivariate statistical method of cluster analysis [24]. The essence of cluster analysis is to collect a sample of similar variables. In cluster analysis, the statistics that reflect the nature of the sample or variable are termed the clustering statistics. Euclidean distance has been used well in similar

research [25, 26]. In this paper, cluster analysis, which is a multivariate statistical method, is performed using SPSS software.

According to the number and characteristics of the samples, the number of clusters is determined. In this paper, after determining the number of clusters, the number of classifications decreases successively, and the classification results are compared. For jars, the number of classifications is 9, 8 and 7. As shown in Table 1, among the three categories, G23:2 belongs to category 1. By observing all kinds of objects, classes with similar characteristics are combined into groups. In the clustering results, a few kinds of objects cannot be combined with other classes, and they can be seen as special individuals. The results are as follows (**Table 1**).

The first group consists of classes 1, 3, 8, and 9 with the following main characteristics: the overall shape is similar to a square, the length of the ear is a third of the height, and the degree of the saddle mouth is obvious.

The second group consists of classes 2 and 5 with the following main characteristics: the overall shape is lanky and square, the length of the ear is a third of the height, and the degree of the saddle mouth is obvious.

The third group consists of classes 4, 6 and 7 with the following main characteristics: the overall shape is lanky, the length of the ear is a third of the height, and the degree of the saddle mouth is not obvious.

Table 1 Clustering results of the jars

NO*: the number of jars

9*: 9 classes, 8*: 8 classes, and 7*: 7 classes

NO	NO*	9*	8*	7*	NO	NO*	9*	8*	7*	NO	NO*	9*	8*	7*
1	G23:2	1	1	1	22	M26:7	2	2	2	43	M57:7	2	2	2
2	G32:2	2	2	2	23	M27:1	2	2	2	44	M57:8	4	4	4
3	G33:1	3	3	3	24	M29:2	1	1	1	45	M58:1	4	4	4
4	G34:5	2	2	2	25	M29:3	4	4	4	46	M58:2	2	2	2
5	G62:3	2	2	2	26	M30:2	4	4	4	47	M59:2	4	4	4
6	G63:5	4	4	4	27	M30:3	6	5	5	48	M60:13	4	4	4
7	H10:4	4	4	4	28	M32:1	2	2	2	49	M61:4	2	2	2
8	H3:1	2	2	2	29	M32:3	4	4	4	50	M61:5	4	4	4
9	H3:2	2	2	2	30	M33:1	4	4	4	51	M61:7	2	2	2
10	H7:4	4	4	4	31	M33:3	2	2	2	52	M61:8	4	4	4
11	M10:3	4	4	4	32	M40:1	6	5	5	53	M62:4	2	2	2
12	M12:2	4	4	4	33	M40:3	2	2	2	54	M64:1	2	2	2
13	M14:6	5	2	2	34	M41:1	4	4	4	55	M64:3	5	2	2
14	M16:2	4	4	4	35	M4:4	2	2	2	56	M65:5	2	2	2
15	M17:1	4	4	4	36	M4:5	4	4	4	57	M6:6	7	6	6
16	M18:2	4	4	4	37	M48:4	7	6	6	58	M7:12	4	4	4

17	M2	1	1	1	38	M49:7	7	6	6	59	M7:13	9	8	7
18	M24:2	2	2	2	39	M51:1	8	7	1	60	M7:9	1	1	1
19	M24:3	2	2	2	40	M51:5	4	4	4	61	M8:1	2	2	2
20	M25:11	2	2	2	41	M53:1	2	2	2	62	T662:2	2	2	2
21	M25:8	4	4	4	42	M54:5	4	4	4					

B. Elliptic Fourier descriptors

Elliptical Fourier descriptors (EFDs), originally proposed by Kuhl and Giardina [8], can delineate any type of shape with a closed two-dimensional contour. The two-dimensional contours are analyzed by the elliptical Fourier descriptors, and the steps are as follows. In this paper, the calculations are achieved through MATLAB programming. In MATLAB, we normalized all formulas and data.

Step 1. The maximum number of harmonics is determined by the elliptical Fourier series formula. The elliptical Fourier descriptor of the two-dimensional contour is calculated through the maximum number of harmonics. The procedure of the elliptical Fourier series approximation involves the transformation of the (x, y) coordinate points on the curve in two dimensions into a pair of equations written as a function of a third variable (t). The Fourier coefficients are then calculated based on a discrete Fourier series approximation of the chain-code boundary contours. The elliptical Fourier series approximation of a closed contour projected on the x- and y-axes can be defined as follows:

$$X_N(t) = A_0 + \sum_{n=1}^N a_n \cos\left(\frac{2n\pi t}{T}\right) + b_n \sin\left(\frac{2n\pi t}{T}\right) \quad (1)$$

$$Y_N(t) = C_0 + \sum_{n=1}^N c_n \cos\left(\frac{2n\pi t}{T}\right) + d_n \sin\left(\frac{2n\pi t}{T}\right)$$

where t is the steps required to move a unit pixel along the closed contour such that $t_{p-1} < t < t_p$ for $1 \leq p \leq K$, N is the number of Fourier harmonics, and K is the total number of chain-coded points. A_0 and C_0 are coefficients corresponding to frequency 0. These coefficients define the mean size of the contour. If the contour between the (i-1)-th and the i-th chain-coded points is linearly interpolated and the length of the contour from the starting point to the p-th point and the perimeter of the contour are denoted t_p and T, respectively, then

$$t_y = \sum_{i=1}^p \Delta t_i \quad (2)$$

T is the basic period of the chain code, which is the overall step used to traverse the entire contour, $T = K \Delta t_i$, where Δt_i is the distance between the (i-1)th and ith points. The Kth point is denoted by x_p . Then,

$$x_p = \sum_{i=1}^p \Delta x_i, \text{ and } y_p = \sum_{i=1}^p \Delta y_i \quad (3)$$

where Δx_i and Δy_i are the distances along the x- and y-axes between the (i-1)th and ith points, respectively. Assuming linear interpolation between the neighboring points, the EFDs in Eq. (1) of the n-th harmonic (a_n , b_n , c_n and d_n) can be calculated using the following equations:

$$a_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta x_p}{\Delta t_p} \left(\cos\left(\frac{2n\pi t_p}{T}\right) - \cos\left(\frac{2n\pi t_{p-1}}{T}\right) \right)$$

$$b_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta x_p}{\Delta t_p} \left(\sin\left(\frac{2n\pi t_p}{T}\right) - \sin\left(\frac{2n\pi t_{p-1}}{T}\right) \right) \quad (4)$$

$$c_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta y_p}{\Delta t_p} \left(\cos\left(\frac{2n\pi t_p}{T}\right) - \cos\left(\frac{2n\pi t_{p-1}}{T}\right) \right)$$

$$d_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta y_p}{\Delta t_p} \left(\sin\left(\frac{2n\pi t_p}{T}\right) - \sin\left(\frac{2n\pi t_{p-1}}{T}\right) \right)$$

The number of harmonics required is estimated using the average Fourier power spectrum. The Fourier power of a harmonic is proportional to the amplitude and provides a measure of the amount of shape information described by the following equation.

$$\text{Fourier power} = \frac{\sum_{n=1}^N (a_n^2 + b_n^2 + c_n^2 + d_n^2)}{2} \quad (5)$$

In this case, the Fourier harmonics are truncated at the value $N=14$, at which point the average cumulative power is 99% or more of the total average power. This is calculated with N_{\max} , which is equal to half the number of boundary points. The boundary contour detection and Fourier series approximation with a Fourier power of 99% is presented in **Fig. 5**.

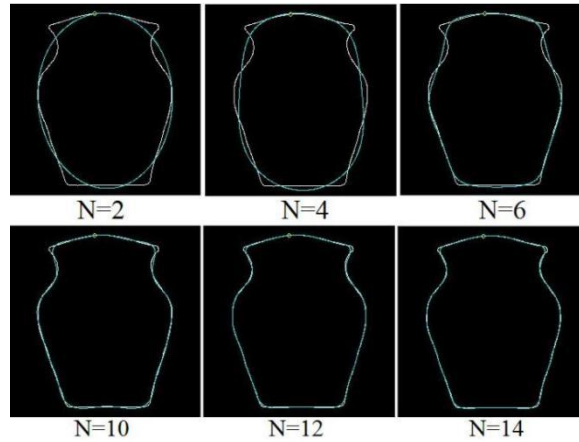


Fig. 5 Contour lines of the jar approximations with different maximum numbers of

harmonics

Step 2. According to the elliptic Fourier descriptor of the two-dimensional contour, principal component analysis is carried out to determine the number where the cumulative contribution rate of the principal component is greater than 90%.

The principal component analysis is calculated according to the elliptical Fourier coefficients of the pottery model. Based on the extraction method of the elliptical Fourier descriptor, this analysis can obtain a vector for the elliptical Fourier descriptor such that when $N_{\max}=14$, the vector is $[a_1, b_1, c_1, d_1, \dots, a_{20}, b_{20}, c_{20}, d_{20}]$. All the samples of the elliptical Fourier descriptor vector can be expressed as

$$G_i = [a_{i1}, a_{i2}, a_{i3}, a_{i4}, \dots, a_{in}, b_{in}, c_{in}, d_{in}],$$

where i is the number of samples, $i=1, 2, \dots, p$, and n is the maximum number of harmonics. The matrix of the Fourier descriptor can be expressed as

$$F = \begin{pmatrix} G_1 \\ \vdots \\ G_p \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pn} \end{pmatrix} \quad (6)$$

The principal component analysis using Eq. 6 is calculated as the cumulative contribution rate of the principal component and the component score coefficient matrix. When the cumulative contribution rate of the principal component reaches 90%, the first principal component can represent the characteristics of the contour lines and reduce the dimensionality.

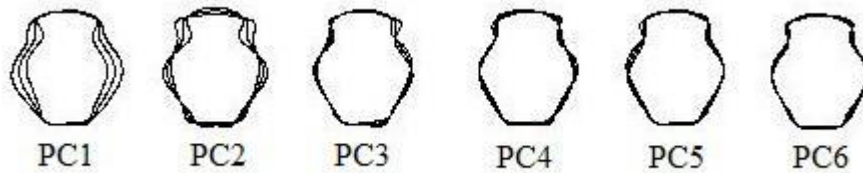


Fig. 6. Variation in the contour lines of the first six principal components

Step 3. The principal component scores are calculated based on the maximum number of harmonics and the cumulative contribution rate of the principal components.

Table 2 Principal component scores according to the first ten principal components

NO.	PC1	PC2	PC3	PC4	PC5	PC6
1.G23: 2	0.14951	-0.02961	0.03450	0.01528	0.02022	0.00601
2.G32: 2	0.05740	0.01565	-0.02334	0.02262	-0.01983	0.00201
3.G33: 1	0.14465	0.06596	0.01210	0.00567	0.00990	-0.01626
4.G34: 5	-0.00806	-0.04752	0.00539	0.01842	-0.01880	0.01020
.....
59.M7: 13	0.13532	-0.05195	0.06982	-0.09018	-0.00608	-0.02102
60.M7: 9	0.10225	0.03736	-0.00231	0.01709	0.02631	-0.01304
61.M8: 1	-0.04005	0.02327	0.01517	0.00128	0.00222	0.00743

Step 4. The principal component scores are analyzed by k-means clustering. Finally, the classification results are obtained.

Based on the principal component scores, the pottery basin shapes are classified by the K-means clustering method. The Euclidean distance is usually chosen as the similarity measure in the K-means clustering algorithm, which usually relates to all attributes. The principal component scores were processed in the K-means clustering analysis using SPSS software. According to the k-means clustering results, the main features of all kinds of objects are as follows.

The first group: The overall shape is dumpy with a round shoulder, and the degree of decrease is not obvious from the abdominal diameter to the bottom diameter.

The second group: The overall shape is close to a square with an angular shoulder, and the degree of decrease is obvious from the abdominal diameter to the bottom diameter.

The third group: The overall shape is lanky with an angular shoulder, and the degree of decrease is obvious from the abdominal diameter to the bottom diameter.

The fourth group: The overall shape is lanky with a round shoulder, and the degree of decrease is not obvious from the abdominal diameter to the bottom diameter.

Table 3 K-means clustering results of jars

NO*: the number of jars

C*: Clustering Results

E*: Euclidean distance

NO	NO*	C*	E*	NO	NO*	C*	E*	NO	NO*	C*	E*
1	G23: 2	2	0.09	22	M26: 7	4	0.11	43	M57: 7	2	0.09
2	G32: 2	2	0.07	23	M27: 1	4	0.10	44	M57: 8	3	0.07
3	G33: 1	2	0.06	24	M29: 2	2	0.07	45	M58: 1	3	0.05
4	G34: 5	4	0.06	25	M29: 3	4	0.08	46	M58: 2	4	0.09
5	G62: 3	4	0.08	26	M30: 2	4	0.06	47	M59: 2	3	0.07
6	G63: 5	4	0.00	27	M30: 3	4	0.11	48	M6: 6	4	0.09
7	H10: 4	4	0.08	28	M32: 1	4	0.07	49	M60: 13	3	0.05
8	H3: 1	2	0.08	29	M32: 3	4	0.08	50	M61: 4	3	0.07
9	H3: 2	2	0.09	30	M33: 1	3	0.08	51	M61: 5	3	0.07
10	H7: 4	2	0.11	31	M33: 3	4	0.09	52	M61: 7	4	0.06
11	M10: 3	3	0.12	32	M4: 4	2	0.08	53	M61: 8	3	0.04
12	M12: 2	3	0.03	33	M4: 5	3	0.09	54	M62: 4	2	0.10
13	M14: 6	4	0.06	34	M40: 1	4	0.08	55	M64: 1	3	0.08
14	M16: 2	4	0.08	35	M40: 3	2	0.10	56	M64: 3	4	0.08
15	M17: 1	4	0.08	36	M41: 1	3	0.00	57	M65: 5	3	0.09
16	M18: 2	3	0.09	37	M48: 4	2	0.10	58	M7: 12	3	0.08
17	M2	2	0.09	38	M49: 7	3	0.08	59	M7: 13	1	0.00

18	M24: 2	2	0.10	39	M51: 1	2	0.12	60	M7: 9	2	0.00
19	M24: 3	4	0.04	40	M51: 5	2	0.08	61	M8: 1	3	0.06
20	M25: 11	2	0.10	41	M53: 1	3	0.09	62	T662: 2	4	0.05
21	M25: 8	3	0.09	42	M54: 5	3	0.10				

Result and discussion

Based on a quantitative analysis of one-dimensional data and the elliptic Fourier descriptors of two-dimensional data, the different characteristics of jars are shown in the classification results. In the clustering analysis of the ratios, the overall shape and the degree of the saddle mouth are obvious, while the difference in the length of the ear is not obvious. In the k-means clustering, the overall shape, the curve of the shoulders and the shape of the abdomen are obvious. Based on the above factors, the final classification results can be compared with the results of the clustering analysis of the ratios and k-means clustering. Considering that the characteristics of pottery are distinguished by two classification methods, the jars are classified according to four parameters: the overall shape, the shoulder shape, the lower abdomen shape and the mouth shape. Specifically, in terms of the overall shape, the jars are divided into square bodies (height \approx abdominal diameter) or thin bodies (height $>$ abdominal diameter). In terms of the shoulder shape, the jars can be divided into round shoulders (the curvature of the shoulders changes more slowly) or folded shoulders (the curvature of the shoulders changes quickly). In terms of the shape of the lower abdomen, the jars are divided into two categories: the contraction of the lower abdomen is not obvious (the abdominal diameter is slightly larger than the bottom diameter), and the contraction of the lower abdomen is obvious (the abdominal diameter is considerably larger than the bottom diameter). In terms of the shape of the mouth, the change in the saddle mouth can be small [the highest height is slightly higher than the shortest height] or the change in the saddle mouth can be large [the highest height is considerably higher than the shortest height]. The final classification result is mutually determined by the clustering analysis and k-means clustering.

Table 4 Results of the classification of the jars

NO*: the number of jars

ONE*: the clustering analysis of the ratios

TWO*: the k-means clustering

NO	NO*	ONE*	TWO*	NO	NO*	ONE*	TWO*	NO	NO*	ONE*	TWO*
1	G23:2	1	2	22	M26:7	2	4	43	M57:7	2	2
2	G32:2	2	2	23	M27:1	2	4	44	M57:8	4	3
3	G33:1	3	2	24	M29:2	1	2	45	M58:1	4	3
4	G34:5	2	4	25	M29:3	4	4	46	M58:2	2	4
5	G62:3	2	4	26	M30:2	4	4	47	M59:2	4	3
6	G63:5	4	4	27	M30:3	6	4	48	M60:13	4	3
7	H10:4	4	4	28	M32:1	2	4	49	M61:4	2	3

8	H3:1	2	2	29	M32:3	4	4	50	M61:5	4	3
9	H3:2	2	2	30	M33:1	4	3	51	M61:7	2	4
10	H7:4	4	2	31	M33:3	2	4	52	M61:8	4	3
11	M10:3	4	3	32	M40:1	6	4	53	M62:4	2	2
12	M12:2	4	3	33	M40:3	2	2	54	M64:1	2	3
13	M14:6	5	4	34	M41:1	4	3	55	M64:3	5	4
14	M16:2	4	4	35	M4:4	2	2	56	M65:5	2	3
15	M17:1	4	4	36	M4:5	4	3	57	M6:6	7	4
16	M18:2	4	3	37	M48:4	7	2	58	M7:12	4	3
17	M2	1	2	38	M49:7	7	3	59	M7:13	9	1
18	M24:2	2	2	39	M51:1	8	2	60	M7:9	1	2
19	M24:3	2	4	40	M51:5	4	2	61	M8:1	2	3
20	M25:11	2	2	41	M53:1	2	3	62	T662:2	2	4
21	M25:8	4	3	42	M54:5	4	3				

After a comprehensive comparison of the classification results, the overall shape of the sample is divided into square bodies and thin bodies. Square bodies are categories 1 and 2 in the clustering analysis and categories 1 and 2 in the k-means clustering. Thin bodies are categories 2 and 3 in the clustering analysis and categories 3 and 4 in the k-means clustering. In terms of the shoulder shape, round shoulders are categories 1 and 4 in the k-means clustering and folded shoulders are categories 2 and 3 in the k-means clustering. In terms of the shape of the lower abdomen, when the contraction of the lower abdomen is not obvious, the jars are classified into categories 1 and 4 in the k-means clustering, and when the contraction of the lower abdomen is obvious, the jars are classified into categories 2 and 3 in the k-means clustering. In terms of the shape of the mouth, when the change in the saddle mouth is small, the jars are classified in category 3 in the clustering analysis, and when the change in the saddle mouth is large, the jars are classified into categories 1 and 2 in the clustering analysis. Individual or small sample categories are incorporated into similar categories according to their morphological characteristics.

Table 5

Characteristic	Classes							
	I	II	III	IV	V	VI	VII*
Square body	√	√				√	√	
Thin body			√	√	√			
Round shoulder	√		√					
Folded shoulder		√		√	√	√	√	
The contraction of the lower abdomen is not obvious	√		√			√		
The contraction of the lower abdomen is obvious		√		√	√		√	
The change in the saddle mouth is large		√		√			√	
The change in the saddle mouth is small	√		√		√	√		

According to the set parameters and two-step classification, the final classification results are as follows (Table 4). *The Roman numerals are the number of categories formed by the characteristic parameters. For jars, the number is 16 ($C_2^1 C_2^1 C_2^1 C_2^1$). Class I includes M7:13. Class II includes G232, G32:2, G33:1, H3:1, H3:2, M2, M24:2, M25:11, M29:2, M40:3, M4:4, M51:1, M57:7, M62:4 and M7:9. Class III includes G34:5, G62:3, G63:5, H10:4, M14:6, M16:2, M17:1, M25:8, M26:7, M27:1, M29:3, M30:2, M30:3, M32:3, M40:1, M4:5, M58:2, M64:1, M64:3 and T662:2. Class IV includes M12:2, M18:2, M24:3, M32:1, M33:1, M33:3, M51:5, M53:1, M61:4, M61:7 and M8:1. Class V includes H7:4, M10:3, M41:1, M48:4, M49:7, M54:5, M57:8, M58:1, M59:2, M60:13, M61:5, M61:8, M65:5, M6:6 and M7:12.

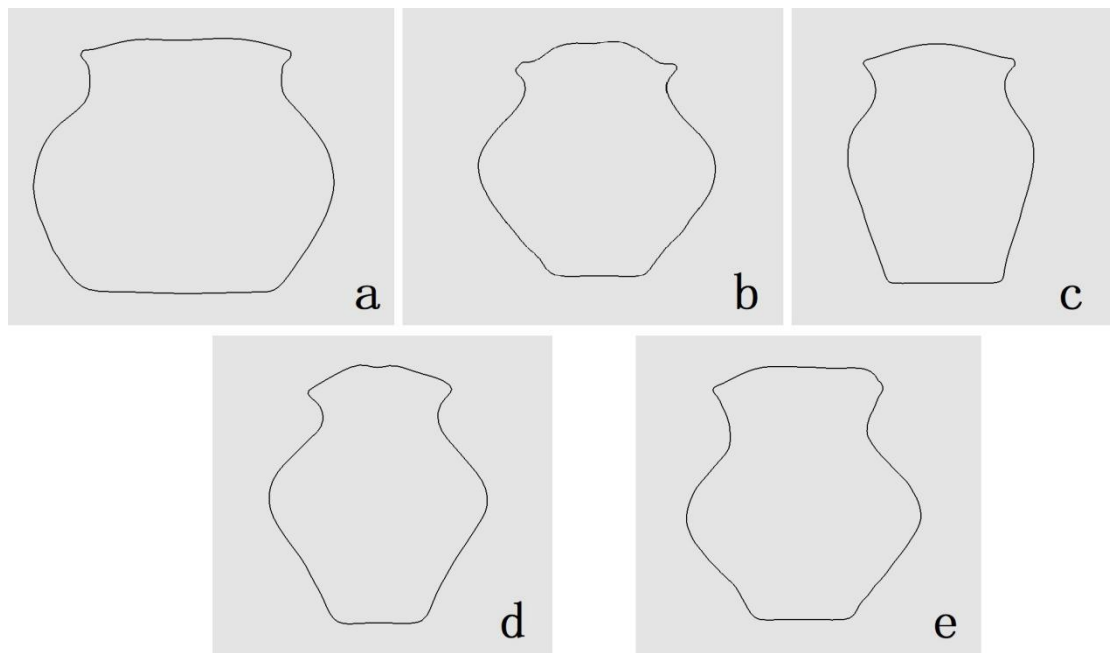


Fig. 7. Representative jars (on the YZ plane in the 3D Builder)

- | | | |
|-------------------|------------------|--------------------|
| a. Class I M7:13 | b. Class II M4:4 | c. Class III M30:2 |
| d. Class IV M61:4 | e. Class V H7:4 | |

Conclusions

With the development of science and technology, 3D scanning technology is being more widely used in archaeological research. Researchers can choose a 3D scanner and postmodel analysis software according to their own research needs. In addition to the research in this paper, a pottery model can be used to create a 3D virtual display, print a 3D model, perform the virtual restoration of cultural relics, and record 3D data. A 3D pottery model is permanent and can be used multiple times as a basic model for further research. In particular, as a result of the significant influence of the COVID-19 pandemic, virtual models will be an important basis for archaeological research in the future. Traditionally, neither archaeological researchers at work nor visitors to museums have avoided prolonged human contact. Archaeologists can conduct a variety of studies based on 3D models at home. Visitors

can visit virtual museums at home. 3D models can be used many times, and the research results can be shared on the Internet.

Due to the mature application of 3D scanning technology, researchers now pay more attention to research content based on 3D scanning and the scientific problems that can be solved by 3D models. In this paper, the focus of our research is to convey a research idea. It is a new exploration and a new attempt to establish an objective classification method for pottery. It provides a new auxiliary tool for traditional archaeological typology research. The computer program in this paper is programmed according to the authors' existing computing power. Researchers can use a variety of methods to extract data based on their computing capabilities. Computer programs are faster and more accurate, and they can avoid multiple touches that are required for manual measurements. The possibility that 3D scanners will offer a renewal and development function in the future means that additional pottery features can be scanned and recorded, such as decorative designs and thicknesses. This capability will increase the number of characteristic factors available for pottery classification.

Traditionally, the accurate classification of pottery shapes largely depended on the experience and knowledge of archaeologists. The mathematical method is an objective method that differs from the methods used in traditional archaeology. It is a tool used to illustrate the relationship between objects, and it enables us to study typology more objectively while avoiding the influence of subjective factors introduced by different researchers. The pottery of the Gansu-Zhanqi site has not been analyzed through traditional archaeological methods. The results of this pottery typology are the first to be published for the site. Therefore, the results of our study provide a reference for traditional archaeology. They contribute to the study of the combination of quantitative and qualitative research in archaeological typology. However, it should be noted that although this method can explain the results of the analysis, it cannot furnish conclusions for archaeological research. Over the past few years, researchers have applied mathematical methods to various areas of traditional archaeology. However, the combination of the two mathematical methods is relatively rare. In this paper, we analyze two-dimensional data—curvature data—and combine them with one-dimensional data. In the present paper, we analyze the overall 3D data of the pottery model through mathematical methods. We hope that these results will form a series that can provide a reference for traditional archaeological research. This makes the results more accurate and valuable.

Supplementary information

Acknowledgements

This work was supported by the University of Science and Technology Beijing, and the Gansu Provincial Institute of Archaeology. The authors would like to thank Guoke Chen and Yishi Yang, who work at the Gansu Provincial Institute of Archaeology.

Authors' contributions

JW performed the data analysis and was a major contributor in writing the

manuscript. WQ and GC provided the archaeological context and pottery samples. All authors read and approved the final manuscript.

Funding

This work was funded by the Shanxi university. This support is gratefully acknowledged.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

References

1. Halekoh U, Vach W. A Bayesian approach to seriation problems in archaeology. *Computational Statistics and Data Analysis*. 2004; 45: 651-673.
2. Donais MK, Wojtas S, Desmond A, Duncan B, George D B. Differentiation of Hypocaust and Floor Tiles at Coriglia, Castel Viscardo (Umbria, Italy) Using Principal Component Analysis (PCA) and Portable X-ray Fluorescence (XRF) Spectrometry. *Applied Spectroscopy*. 2012; 66: 1005-1012.
3. Hodson FR. Cluster analysis and archaeology: some new developments and applications. *World Archaeology*. 1970; 1: 299-320.
4. Otárola-Castillo E. Differences between NISP and MNE in cutmark analysis of highly fragmented faunal assemblages. *Journal of Archaeological Science*. 2009; 37: 1-12.
5. Roger NR. Correspondence Analysis and West Mexico Archaeology: Ceramics from the Long-Glassow Collection. *Bulletin of Latin American Research*. 2016; 35: 384-395.
6. Camiz S, Rova E, Tulli V. Quantitative study of images in Archaeology: II. Symbolic Coding. *Between Data Science and Applied Data Analysis*. Springer Berlin, Heidelberg. 2003; 7: 633-642.
7. Chen TM. Concise archaeological statistics. Science Press; 2013 **(in Chinese)**.
8. Kuhl FP, Giardina CR. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing*. 1982; 18: 236-258.
9. Demir B, Sayinci B, Çetin N, Yaman M, Çömlek R. Shape Discrimination of Almond Cultivars by Elliptic Fourier Descriptors. *Erwerbs-Obstbau* . 2019; 61: 245 – 256.
10. Xia M. Shape analysis of bust slice using Elliptic Fourier. *Journal of Textile Research*. 2014; 35: 107-112 **(in Chinese)**.
11. Wu G, Li D, Hu PP, Zhang YQ. Foot shape prediction using elliptical Fourier analysis. *Textile Research Journal*. 2018; 88: 1026-1037.
12. Fox AN. A study of Late Woodland projectile point typology in New York using elliptical Fourier outline analysis. *Journal of Archaeological Science:Reports*. 2015; 4: 501-509.
13. Gansu provincial institute of archaeology. The bulletin of the Zhanqi site Siwa

culture site in Min county Gansu. Archaeological and cultural relics. 2012; 4:35-47 **(in Chinese)**.

14. Yang YS. The cemetery research of the Zhanqi site Siwa culture site in Min county. Northwestern university master's thesis; 2014 **(in Chinese)**.

15. Wang J. Quantitative analysis of pottery from the Tianma-Qucun Site based on 3D scanning and computer technology. Archaeological and Anthropological Sciences. 2019; 7: 5645 – 5656.

16. The supplier of Creaform Go!SCAN 20 is AMETEK.

17. The supplier of Vxelements software is AMETEK.

18. The supplier of Autodesk meshmixer is Autodesk.

19. The supplier of 3D builder is Microsoft Corporation.

20. Spelitz S, Almeida VM, Auinger CL. Automatic geometry, metrology, and visualisation techniques for 3D scanned vessels. Digital Applications in Archaeology and Cultural Heritage. 2020; 6: 39-50.

21. Rasheed NR, Nordin JM. Reconstruction algorithm for archaeological fragments using slope features. ETRI Journal. 2020; 6: 420-432.

22. Măruțoiu C, Bratu I, Țiplic MI, Măruțoiu VC, Nemeș OF, Neamțu C, Hernanz A. FTIR analysis and 3D restoration of Transylvanian popular pottery from the XVI-XVIII centuries. Journal of Archaeological Science: Reports. 2018; 19: 148-154.

23. Angelo LD, Stefano PD, Pane C. Automatic dimensional characterisation of pottery. Journal of Cultural Heritage. 2017; 26: 118-128.

24. Feng B. Statistical principles. Peking University Press; 2017 **(in Chinese)**.

25. Teng MY. Practice and thinking of mathematical method in the study of archaeological typology. Research of the frontier archaeology (2). Beijing: science press; 2004 **(in Chinese)**.

26. Lang JY. Quantitative Analysis on Potteries unearthed from the Graveyard of Ba Clan at Lijiaba Site, Sichuan university master's thesis; 2007 **(in Chinese)**.