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 (✉ wjustb@163.com)

Shanxi University

wei Qian

University of science and technology Beijing

guoke Chen

Gansu provincial institute of archaeology

Research article

Keywords: pottery, Gansu-Zhanqi site, quantitative analysis, elliptic Fourier descriptor, computer programming, 3D scanning

Posted Date: September 9th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-41185/v5>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Archaeological Science: Reports on April 1st, 2021. See the published version at <https://doi.org/10.1016/j.jasrep.2021.102897>.

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 (3D) models of the pottery were obtained using 3D scanning, and a computer-assisted pottery typology was studied through quantitative analysis and elliptic Fourier descriptor. This method, which can enhance and supplement the traditional methods of classifying pottery in archaeology and thereby enrich the parameters and breadth of pottery analysis, 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, and the classification pattern of each object is an important factor used to determine the type and age of an archaeological culture [1]. Over the past few decades, some archaeologists have utilized mathematical methods such as seriation [2], principal component analysis [3], cluster analysis [4, 5], discriminant analysis [6], multivariate statistics [7], correspondence analysis [8], and principal coordinate analysis [9]. Notably, Chinese archaeologists have used cluster analysis, which is a multivariate statistical method, to classify pottery ware [10].

Traditionally, these mathematical methods have been used alone. The data have

been measured manually. In this paper, we use a computer program to extract data from three-dimensional (3D) pottery models and use the data for quantitative analysis and elliptical Fourier descriptors. As quantitative analysis, Chinese archaeologists have used many times [11, 12, 13]. It has been verified to be correct according to traditional archaeological research. As elliptic Fourier descriptors (EFDs), originally proposed by Kuhl and Giardina [14], can delineate any type of shape with a closed two-dimensional (2D) contour. Over the past several years, researchers have applied elliptical Fourier descriptors to various fields, such as agriculture [15], the human body [16, 17], and ecology [18], but it has rarely been used in archaeological research. This study, however, combines the results of quantitative analysis and elliptical Fourier descriptors to study pottery typology. The final analysis provides additional parameters, which make the final results more accurate.

In this article, we use unearthed pottery from the Gansu-Zhanqi burial site as examples. The Zhanqi site is a part of the Siwa cultural relics (14 BC to 11 BC) along the Tao River in Min County, Gansu Province. 66 tombs were found at the site, and an additional 20 sites were found that included houses, cooking pits, ash pits and sacrificial remains. The unearthed objects included pottery, bronze ware, stone tools, bone artifacts and ornaments. The layers and times of the tombs are clear, and unearthed artifacts are abundant. The Gansu-Zhanqi site is of great significance for the further study of the Siwa culture as considerable amounts of pottery were unearthed at the site [19] (**Fig. 1**), including jars, lis, basins, dous, and bottles, albeit the majority of the pottery consists of jars. According to the shape of the mouth, the jars are divided into flat and saddle jars [20]. In this article, 62 saddle jar examples, which are representative pottery from the Gansu-Zhanqi site, were scanned.

First, it is important to note that the pottery model, which is created using 3D scanning, can be used to create a 3D virtual display, print a 3D model, perform a virtual restoration of cultural relics, and record 3D data. Moreover, 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, however. Second, a quantitative analysis of pottery has been applied to many archaeological studies. Compared with traditional archaeological methods, the cluster analysis results were basically identical to the results provided by the original reports, and the results of this study have been confirmed in the Tianma-Qucun site [13]. While contours can be analyzed in a variety of ways, the elliptic Fourier descriptors is a highly suitable method for this paper. The pottery of the Gansu-Zhanqi site have not been analyzed through traditional archaeological typology, the results of this pottery classification are the first to be published for this site.

Fig. 1. The unearthed pottery^[19]

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

*Li is a ceramic cooking vessel used in ancient China. It appeared in the late Neolithic Age and continued to be popular in the Shang and Zhou dynasties. Most have an open mouth, a round abdomen and three baggy feet.

Unifying the Data Extracted from the Pottery Model

The 3D models were acquired using a Creaform Go!SCAN 20 scanner [21]. 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. Therefore this scanner is more suitable for small cultural relics. VXelements [22], Autodesk Meshmixer [23], and 3D Builder [24] are used for the postprocessing stage of the 3D-scanning process (Fig. 2). It is important to note that since the research on pottery shapes requires less accuracy than does 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 lengths are based on the XZ plane in the 3D Builder.
- d. The contours are based on the YZ plane in the 3D Builder.

The data are extracted from the 3D pottery model and include two kinds of data. While the lengths are used for quantitative analysis, the contours are used for elliptic Fourier descriptors. Over the past few years, scientists have introduced several data extraction methods for 3D pottery models, including Spelitz [25], Rasheed [26], Măruțoiu [27], and Angelo [28]. In this paper, the data are extracted using a computer program in C language (**Fig. 3**). It is further noted that the jars at the Gansu-Zhanqi site are binaural, i.e., have two ears, which is a representative characteristic of pottery from northwest China. Considering the methods and functions of this article, the extraction and analysis of the lengths are based on the XZ plane in the model, which shows the different calibers and heights of the ears, whereas the extraction and analysis of the contours are based on the YZ plane in the model, which shows variations 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 the output.

Methods

A. Quantitative analysis

Step 1. The first step is to obtain the original data. The measurement areas are presented in **Fig. 4**. To a certain extent, these areas represent the characteristics of the jar, which are the data used for the quantitative analysis.

Fig. 4. Measurement areas for a saddle 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.

62 saddle jars were analyzed by examining the overall structure of the pottery model, analyzing the characteristic lengths, and selecting meaningful ratios for the following:

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

Step 3. The third step is to conduct a quantitative analysis using the ratios obtained in the second step. In this paper, we use the multivariate statistical method of cluster analysis [29]. Cluster analysis is a branch of multivariate analysis which is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). The essence of cluster analysis is to collect a sample of similar variables. It is one of important techniques in data mining and exploratory data analysis. The similarity distance between samples was calculated using Euclidean square distance. Euclidean distance is used to calculate the similarity distance between samples which has been used well in similar research [11, 12]. A similarity matrix is formed by calculating the distance between each sample. In the clustering process, the between-class distance is determined by the single linkage method, the complete linkage method, the median method, the average linkage method, the Ward's method and so on. The Ward's method is used in this paper. The similarity distance and the Ward's method were performed using SPSS software.

According to the number and characteristics of the samples, the number of clusters is determined. In this paper, The number of clusters is based on published studies which have similar pottery [11, 12, 13]. It has been verified to be correct according to traditional archaeological research. For jars, the number of classifications is 9. The cluster analysis results and dendrogram are shown in Table S1 and Figure S1. 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 thus, they are regarded as special individuals. The results are as follows.

The group 1 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 group 2 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 group 3 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.

B. Elliptic Fourier descriptors

EFDs, originally proposed by Kuhl and Giardina [8], delineate any type of shape with a closed 2D contour, and the contours are then analyzed using the EFDs. The steps are as follows. In this paper, the calculations are achieved using SHAPE [30].

Step 1. The maximum number of harmonics is determined by the elliptical Fourier series formula. The EFD of the 2D contour is calculated based on 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 is 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 and 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 by 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; $T = tk$, where Δt_i is the distance between the (i-1)th and i-th 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 i-th 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:

$$\begin{aligned} a_n &= \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta x_p}{\Delta t_p} \left(\cos\left(\frac{2n\pi t_p}{T}\right) - \cos\left(\frac{2n\pi t_{p-1}}{T}\right) \right) \\ b_n &= \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta x_p}{\Delta t_p} \left(\sin\left(\frac{2n\pi t_p}{T}\right) - \sin\left(\frac{2n\pi t_{p-1}}{T}\right) \right) \\ c_n &= \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta y_p}{\Delta t_p} \left(\cos\left(\frac{2n\pi t_p}{T}\right) - \cos\left(\frac{2n\pi t_{p-1}}{T}\right) \right) \\ d_n &= \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta y_p}{\Delta t_p} \left(\sin\left(\frac{2n\pi t_p}{T}\right) - \sin\left(\frac{2n\pi t_{p-1}}{T}\right) \right) \end{aligned} \quad (4)$$

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 as 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 90% or more of the total average power. This is calculated as $N_{\max}$, which is equal to half the number of boundary points. The boundary contour detection and the Fourier series approximation with a Fourier power of 90% is presented in **Fig. 5**. According to the maximum harmonic numbers $N=14$, the Elliptic Fourier coefficients were determined using Eqs. (1-4), presented in **Table S2**.

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

Step 2. Principal component analysis. The principal component analysis is calculated according to the elliptical Fourier coefficients of the pottery model. Based on the extraction method of the EFD, this analysis yields a vector for the EFD 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 EFD vector can be expressed as

$$G_i = [ai_1, ai_2, ai_3, ai_4, \dots, a_{in}, b_{in}, c_{in}, d_{in}],$$

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 is expressed as

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

(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 thereby reduce dimensionality. Finally, the principal component scores (**Table S3**) are calculated based on the maximum number of harmonics ($N_{\max}=14$) and the cumulative contribution rate of the principal components (**Fig. S2**).

Step 3. K-means clustering. Based on the principal component scores, the

pottery basin shapes are classified using the k-means clustering method. The Euclidean distance is usually selected as the similarity measure in the k-means clustering algorithm, which usually relates to all attributes. The principal component scores are then processed in the k-means clustering analysis using SPSS software. According to the k-means clustering results (**Table S4**), the main features of all kinds of objects are as follows.

The class □ with the following main characteristics: 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 class □ with the following main characteristics: The overall shape closely resembles a square with an angular shoulder, and the degree of decrease is obvious from the abdominal diameter to the bottom diameter.

The class □ with the following main characteristics: 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 class □ with the following main characteristics: 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.

Results and Discussion

Based on quantitative analysis and Elliptic Fourier descriptors, the different characteristics of jars are presented 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 jars are classified according to four parameters, specifically, the overall shape, the shoulder shape, the lower abdomen shape and the mouth shape. Each parameter is composed of two features, as follows: 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 are divided into round shoulders (the curvature of the shoulders changes more slowly) or folded shoulders (the curvature of the shoulders changes quickly). With respect to the shape of the lower abdomen, the contraction of the lower abdomen is not obvious, i.e., the abdominal diameter is slightly larger than the bottom diameter, or the contraction of the lower abdomen is obvious, i.e., the abdominal diameter is considerably larger than the bottom diameter. With respect to the shape of the mouth, the change in the saddle mouth may be small, i.e., the highest height is slightly higher than the shortest height, or the change in the saddle mouth can be large, i.e., the highest height is considerably higher than the shortest height.

For the four parameters, the mouth shape is analyzed according to the clustering analysis of the ratios, the shoulder shape and the lower abdomen are analyzed according to the k-means clustering. The three parameters are independent of each other. The overall shape is analyzed according to the clustering analysis of the ratios and the k-means clustering. The classification results can be compared with the results of the clustering analysis of the ratios and the k-means clustering. For example, in the clustering analysis of the ratios, the group 2 consists of classes 2 and 5 with the following main characteristics: the overall shape is lanky and square. In the k-means clustering, the class □ is square, the class □ is lanky. In the table 1, G32:2 is Classes (HCA) 2 and Classes (k-means) □ which is belong to square body. G34:5 is Classes (HCA) 2 and Classes (k-means) ④ which is belong to thin body. It's important to note that a few kinds of objects cannot be combined with other classes, and thus, they are regarded as special individuals. The final classification result is mutually determined by the clustering analysis and k-means clustering (**Table 1**).

Table 1 Results of the classification of the jars

NO	The number of jars	Classes (HCA)	Classes (k-means)	NO	The number of jars	Classes (HCA)	Classes (k-means)
1	G23:2	1	②	32	M40:1	6	④
2	G32:2	2	②	33	M40:3	2	②
3	G33:1	3	②	34	M41:1	4	③
4	G34:5	2	④	35	M4:4	2	②
5	G62:3	2	④	36	M4:5	4	③
6	G63:5	4	④	37	M48:4	7	②
7	H10:4	4	④	38	M49:7	7	③
8	H3:1	2	②	39	M51:1	8	②
9	H3:2	2	②	40	M51:5	4	②
10	H7:4	4	②	41	M53:1	2	③
11	M10:3	4	③	42	M54:5	4	③
12	M12:2	4	③	43	M57:7	2	②
13	M14:6	5	④	44	M57:8	4	③
14	M16:2	4	④	45	M58:1	4	③
15	M17:1	4	④	46	M58:2	2	④
16	M18:2	4	③	47	M59:2	4	③
17	M2	1	②	48	M60:13	4	③
18	M24:2	2	②	49	M61:4	2	③
19	M24:3	2	④	50	M61:5	4	③

20	M25:11	2	②	51	M61:7	2	④
21	M25:8	4	③	52	M61:8	4	③
22	M26:7	2	④	53	M62:4	2	②
23	M27:1	2	④	54	M64:1	2	③
24	M29:2	1	②	55	M64:3	5	④
25	M29:3	4	④	56	M65:5	2	③
26	M30:2	4	④	57	M6:6	7	④
27	M30:3	6	④	58	M7:12	4	③
28	M32:1	2	④	59	M7:13	9	①
29	M32:3	4	④	60	M7:9	1	②
30	M33:1	4	③	61	M8:1	2	③
31	M33:3	2	④	62	T662:2	2	④

Following a comprehensive comparison of the classification results, the overall shape of the sample is divided into square bodies and thin bodies. Square bodies are included in groups 1 (classes 1, 3, 8, and 9) and 2 (classes 2 and 5) in the clustering analysis of the ratios and classes ① and ② in the k-means clustering. Thin bodies are included in groups 2 (classes 2 and 5) and 3 (classes 4, 6 and 7) in the clustering analysis of the ratios and classes ③ and ④ in the k-means clustering. In terms of the shoulder shape, round shoulders are included classes ① and ④ in the k-means clustering and folded shoulders are included in classes ② and ③ in the k-means clustering. Regarding the shape of the lower abdomen, when the contraction of the lower abdomen is not obvious, the jars are classified into classes ① and ④ in the k-means clustering, and when the contraction of the lower abdomen is obvious, the jars are classified into classes 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 into group 3 (classes 4, 6 and 7) in the clustering analysis of the ratios, and when the change in the saddle mouth is large, the jars are classified into groups 1 (classes 1, 3, 8, and 9) and 2 (classes 2 and 5) in the clustering analysis of the ratios. Individual or small sample classes are incorporated into similar classes according to their morphological characteristics.

There are four parameters, and each parameter is composed of two features, the final number of classes is 16 ($C_2^1 C_2^1 C_2^1 C_2^1$). The Roman numerals denote the number of classes formed by the characteristic parameters (**Fig. S3**). According to the set parameters and the two-step classification, the final classification results are as follows (**Table 2**). It should be noted that in Table 2, there are 16 classes in theory (I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, XIV, XV, XVI). In this article, 62 saddle jar examples, only in 6 out of the 16 classes defined real individuals are included. Therefore, the final classification results have 6 classes as follows: Class I includes M7:13. Class II includes G23:2, 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 G63:5, H10:4, M16:2, M17:1, M29:3, M30:2, M30:3, M32:3, M40:1 and M6:6. Class IV includes M53:1, M61:4, M64:1, M65:5, M8:1, H7:4, M48:4 and M51:5. Class V includes G34:5, G62:3, M14:6, M24:3, M26:7, M27:1, M32:1, M33:3, M58:2, M61:7, M64:3 and T662:2. Class VI includes M10:3, M12:2, M18:2, M25:8, M33:1, M41:1, M4:5, M49:7, M54:5, M57:8, M58:1, M59:2, M60:13, M61:5, M61:8 and M7:12(**Fig. 6**).

Table 2

Characteristic	Classes								
	I	II	III	IV	V	VI	...	XVI	
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	√		√			√

Fig. 6. 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 M24:3 f. Class VI M41:1

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.

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 concept, specifically, a new exploration and a new attempt to establish an objective classification method for pottery. As such this study 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, a capability that increases the number of characteristic factors available for pottery classification.

Traditionally, the accurate classification of pottery shapes largely depended on the experience and knowledge of the archaeologists. The mathematical method, however, 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. As 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, and 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 2D data and combine them with 1D 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, thereby enhancing the accuracy and value of the results.

Figure legends

Fig. 1. The unearthed pottery ^[19]

1.Saddle jar 2.Flat jar 3.Colored pottery 4. Dou 5. Bottle 6.Bottle 7. Flat jar 8. Bottle
9. Li* 10.Li* (*Li is a ceramic cooking vessel used in ancient China. It appeared in the late Neolithic Age and continued to be popular in the Shang and Zhou dynasties. Most have an open mouth, a round abdomen and three baggy feet).

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 lengths are based on the XZ plane in the 3D Builder.
- d. The contours are based on the YZ plane in the 3D Builder.

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 the output.

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

Fig. 5. Contour lines of the jar approximations with different maximum numbers of harmonics**Fig. 6.** 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 M24:3	f. Class VI M41:1

Supplementary information

Additional file 1. Additional tables

Additional file 2. Additional figures

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 and the University of Science and Technology Beijing. Supported by Program for the Philosophy and Social

Sciences Key Research Base of Higher Education Institutions of Shanxi (PSRB) (Grant Number: 20200111).

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. Hein A. The problem of typology in Chinese archaeology. *Early China*. 2016; 2: 1-32.
2. Halekoh U, Vach W. A Bayesian approach to seriation problems in archaeology. *Computational Statistics and Data Analysis*. 2004; 45: 651-673.
3. 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.
4. Hodson FR. Cluster analysis and archaeology: some new developments and applications. *World Archaeology*. 1970; 1: 299-320.
5. Whittaker JC, Caulkins D, Kamp KA. Evaluating Consistency in Typology and Classification. *Journal of Archaeological Method and Theory*. 1998; 2: 129-164.
6. Karasik A, Smilansky U. Computerized morphological classification of ceramics. *Journal of Archaeological Science*. 2011; 38: 2644-2657.
7. Otárola-Castillo E. Differences between NISP and MNE in cutmark analysis of highly fragmented faunal assemblages. *Journal of Archaeological Science*. 2009; 37: 1-12.
8. Roger NR. Correspondence Analysis and West Mexico Archaeology: Ceramics from the Long-Glassow Collection. *Bulletin of Latin American Research*. 2016; 35: 384-395.
9. 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.
10. Chen TM. Concise archaeological statistics. Science Press; 2013 **(in Chinese)**.
11. 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)**.
12. Lang JY. Quantitative Analysis on Potteries unearthed from the Graveyard of Ba Clan at Lijiaba Site, Sichuan university master's thesis; 2007 **(in Chinese)**.
13. 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.
14. Kuhl FP, Giardina CR. Elliptic Fourier features of a closed contour. *Computer*

Graphics and Image Processing. 1982; 18: 236-258.

15. 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.

16. Xia M. Shape analysis of bust slice using Elliptic Fourier. *Journal of Textile Research*. 2014; 35: 107-112 **(in Chinese)**.

17. Wu G, Li D, Hu PP, Zhang YQ. Foot shape prediction using elliptical Fourier analysis. *Textile Research Journal*. 2018; 88: 1026-1037.

18. 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.

19. 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)**.

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

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

22. The supplier of Vxelements software is AMETEK.

23. The supplier of Autodesk meshmixer is Autodesk.

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

25. 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.

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

27. 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.

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

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

30. The supplier of SHAPE is Hiroyoshi Iwata.

Figures

Figure 1

The unearthed pottery [19] 1.Saddle jar 2.Flat jar 3.Colored pottery 4. Dou 5. Bottle 6.Bottle 7. Flat jar 8. Bottle 9. Li* 10.Li* (*Li is a ceramic cooking vessel used in ancient China. It appeared in the late Neolithic Age and continued to be popular in the Shang and Zhou dynasties. Most have an open mouth, a round abdomen and three baggy feet).

Figure 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 lengths are based on the XZ plane in the 3D Builder. d. The contours are based on the YZ plane in the 3D Builder.

Figure 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 the output.

Figure 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

Figure 5

Contour lines of the jar approximations with different maximum numbers of harmonics

Figure 6

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 M24:3 f. Class ∞ M41:1

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryinformation.docx](#)