Partition-Based Sex Determination Method for Damaged Skulls: Postprint

Authors: Yang Wen, Liu Xiaoning, Zhu Fei, Geng Guohua, Zhao Qianna

Date: 2018-06-19T00:00:00+00:00

Abstract

Skull sex identification represents a primary and critical step in identity authentication within forensic anthropology, wherein sex identification from damaged skulls holds greater research value in practical applications. To identify the sex of damaged skulls and narrow the search scope of criminal investigations, this paper proposes a sex discrimination model for incomplete skulls. First, the skull is partitioned into seven regions, with feature points marked and non-measurable features quantified. Then, a forward stepwise regression method based on maximum likelihood estimation is utilized to select the optimal feature subset for each region, and Logistic regression is employed to establish sex discrimination models for the seven regions, which are validated through leave-one-out cross-validation. Finally, final sex discrimination for incomplete female and male skulls is achieved. Experimental results demonstrate that each region can independently determine skull sex, and the discrimination accuracy increases with the number of combined regions.

Full Text

Preamble

Sex Determination of Incomplete Skull Based on Partition

Yang Wen, Liu Xiaoning, Zhu Fei, Geng Guohua, Zhao Qianna
(School of Information Science & Technology, Northwest University, Xi' an 710127, China)

Abstract: Skull gender identification is the first and crucial step in forensic anthropology authentication. In practice, sex identification of incomplete skulls holds greater research value. To identify the gender of damaged skulls and narrow the scope of criminal investigations, this paper proposes an incomplete skull sex discrimination model. First, the skull is divided into seven partitions, feature points are marked, and non-measurable features are quantified. Then, the optimal feature subset for each partition is selected using a forward stepwise regression method based on maximum likelihood estimation. Seven partition-specific gender discrimination models are established through Logistic regression and validated using leave-one-out cross-validation. Finally, the ultimate sex determination for incomplete female and male skulls is achieved. Experimental results demonstrate that each partition can determine skull gender, and the discrimination accuracy increases with more partition combinations.

Keywords: gender identification of skull; partition; stepwise regression; leave-one-out cross-validation

0 Introduction

Skull sex determination holds significant value in forensic medicine, anthropology, facial reconstruction, and identification of unknown remains. With the maturation of computer technology and statistical research theory, leveraging modern tools to complete various studies has become a compelling research focus. Research on skull sex discrimination using three-dimensional digital information carries major scientific and practical value. Analyzing sex through measurement of skull-related indicators has become a hot topic in forensic anthropology.

In recent years, scholars both domestically and internationally have conducted research on skull sex discrimination. Twisha et al. [1] used eight cranial measurements to establish discriminant and logistic regression equations for sexual dimorphism determination, achieving correct identification rates of 92% for males and 80.9% for females. Amores-Ampuero [2] measured six characteristic indicators from 109 Spanish skull samples, establishing a discriminant function with 75.7% accuracy. Ajanovic et al. [3] measured seven linear diameters of Bosnian skulls to establish discriminant functions. Domestic researchers such as Shui Wuyang et al. utilized stepwise Fisher' s method to establish multivariate gender discriminant functions for adult skulls in the Xi' an region, achieving 87.5% accuracy for males and 86.67% for females using four indicators. Li Chunbiao et al. applied Fourier transform and multivariate stepwise discriminant analysis to determine sex for Northeast Chinese adult skulls, with male identification rates of 84.21% and female rates of 83.33%. Li Ming et al. established sex determination equations for Southwest Chinese skulls based on cranial circumference and nasal height measurements, achieving 89.2% accuracy for males and 90.0% for females. All these methods target complete skulls.

However, in practice, situations involving only skull fragments are common, requiring sex identification of these remains. Some researchers have attempted sex determination using single skull bones. For example, Song Hongwei et al. applied discriminant analysis to the maxilla, frontal bone, occipital bone, and parietal bone, demonstrating that single skull bones also exhibit significant sexual differences with relatively high discrimination rates. To address the problem of varying skull fragment conditions in practical applications, this paper partitions the skull and establishes different sex identification models for different partitions, then fuses multiple partition models to construct a damaged skull sex identification model with the highest possible accuracy.

1.1 Feature Point Calibration

Referencing forensic feature point definition standards and craniofacial morphology research literature [8,9], this paper defines 62 skull feature points for sex identification, including 16 midsagittal plane feature points and 46 lateral skull feature points. These points are distributed across the entire skull and can represent both facial feature and contour characteristics.

Using self-developed software, we performed three-dimensional modeling of skulls and aligned them in the Frankfurt coordinate system. We then utilized an in-house feature point labeling system to manually mark the 62 skull feature points. The labeling results are shown in [Figure 1: see original paper].

As evident from the data, sexual differences in skulls are relatively pronounced. Some features can be obtained through measurement methods, such as Euclidean distance, geodesic distance, angles, surface area, and volume. Other features, like brow ridge prominence and external occipital protuberance prominence, cannot be directly measured and thus require quantification of these non-measurable characteristics.

For quantifying external occipital protuberance prominence, we transformed its descriptive assessment into an angle measurement. In the midsagittal plane, the angle is formed between the line connecting the upper and lower rostral protrusion points and the lower occipital bone line. Based on the angle size, as shown in Figure 2: see original paper, we classify it into six grades: absent, slight, moderate, prominent, very prominent, and rostrum-like downward extension, quantified as numerical values 1-6, as illustrated in Figure 2: see original paper.

1.2 Skull Sex Feature Quantification

Sex differences in skull surface features primarily manifest in the morphology of cranial surface anatomical structures [10,11]. For skull regions with pronounced sexual dimorphism, observers can relatively easily determine sex based on experience. lists the differences between male and female skull surface features.

For quantifying brow ridge prominence, the sagittal morphology of the glabellar elevation and nasal root depression approximates a portion of a circle. We use least squares fitting to determine the circle’ s radius. Based on the radius size, brow ridge prominence is classified into six grades: not prominent, slight, moderate, prominent, very prominent, and robust, quantified as numerical values 1-6, as shown in [Figure 3: see original paper].

2.1 Skull Partitioning and Feature Selection

When skulls are fragmented, not only does the difficulty of sex determination increase, but the discrimination accuracy also decreases significantly. Based on physiological structure research by anthropologists, we partition the skull into seven regions: orbital, frontal, zygomatic, maxillary, parietal, occipital, and mandibular regions. We further investigate how to integrate the evaluation indicators from each region to obtain a comprehensive determination result with the highest discrimination rate for damaged skulls.

From 267 three-dimensional digital models of Uyghur skulls, we extracted 43 cranial feature indicators, including linear distances, angles, curved distances, and quantified indices, as shown in . We applied forward stepwise regression based on maximum likelihood estimation to select optimal feature subsets for the seven partitions, as presented in .

2.2 Partition-Based Skull Sex Classification Model

Logistic regression analysis [12,13] predicts the probability of an event occurrence through a multivariate regression relationship between a dependent variable and multiple independent variables. The independent variables in logistic regression can be continuous or discrete and need not satisfy normal distribution, giving it broader applicability compared to other models. For binary classification, it uses a nonlinear S-shaped curve function and employs maximum likelihood estimation to ensure optimal fitting at each point. This paper applies logistic regression to skull sex identification and verifies its accuracy through experimental results.

For a given skull training dataset $\{(x_i, y_i), i = 1, 2, ..., N\}$, where $x_i \in \mathbb{R}^n$ represents the skull model and $y_i \in \{0, 1\}$ indicates whether the skull is male or female. When applying the Logistic regression model, we can use maximum likelihood estimation to estimate the parameters.

Assuming the probability that the i-th skull sample is classified as male is $\pi_i = P(y_i = 1|x_i)$, then the probability of being classified as female is $1-\pi_i = P(y_i = 0|x_i)$. The likelihood function is:

$$L(\alpha, \beta) = \prod_{i=1}^{N} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

The log-likelihood function is:

$$l(\alpha,\beta)=\sum_{i=1}^{N}[y_i\log(\pi_i)+(1-y_i)\log(1-\pi_i)]$$

$$=\sum_{i=1}^{N}[y_i\log(P(y_i=1|x_i))+(1-y_i)\log(P(y_i=0|x_i))]$$

$$=\sum_{i=1}^{N}[y_i(\alpha+\beta x_i)-\log(1+e^{\alpha+\beta x_i})]$$

To find the maximum, we obtain estimates of α and β. This transforms the problem into an optimization problem with the log-likelihood function as the objective. This paper uses the quasi-Newton method for solution. Assuming $\widehat{\alpha}$ and $\widehat{\beta}$ are the maximum likelihood estimates of α and β, the Logistic regression model is expressed as:

$$P(y_i=1|x_i)=\frac{e^{\hat{\alpha}+\hat{\beta}x_i}}{1+e^{\hat{\alpha}+\hat{\beta}x_i}}$$

$$P(y_i=0|x_i)=\frac{1}{1+e^{\hat{\alpha}+\hat{\beta}x_i}}$$

The seven partitions of the skull are represented as $R_i, i=1,2,...,7$. Based on , we determine the optimal cranial feature indicators for the seven partitions and establish sex discrimination models using Logistic regression, denoted as $h_i(R_i)$. We validate the model accuracy a_i^m and a_i^f for male and female skull classification respectively. For the sex determination model of damaged skulls, we have:

$$P(y=1|x)=\sum_{i=1}^{7}a_i^m h_i(R_i)$$

$$P(y=0|x)=\sum_{i=1}^{7}a_i^f h_i(R_i)$$

where $h_i(R_i)$ represents the confidence level of classifying skull sample x as male or female by the model. If $P(y=1|x)>P(y=0|x)$, then x is classified as male; otherwise, x is classified as female. The flowchart of the partition-based skull sex identification model is shown in [Figure 4: see original paper].

3 Experimental Results

To compare the differences among skull partition classifiers and their discrimination accuracy, we employed forward stepwise regression based on maximum likelihood estimation to establish Logistic regression models for the seven regions. We validated these models using leave-one-out cross-validation to obtain classification models and back-substitution test accuracy, as shown in .

3.1 Gender Classification Results for Each Skull Partition

From , we can observe:

a) Among the seven skull partitions, the frontal bone partition achieves the highest sex classification accuracy at 82.0%.

b) The mandibular and occipital partitions rank second with classification accuracies of 81.8% and 81.4% respectively.

c) The maxillary and zygomatic partitions have moderate accuracies of 80.9% and 79.2% respectively.

d) The orbital and parietal partitions have the lowest classification accuracies at 62.9% and 60.9% respectively.

From these results, we can infer that the frontal bone, mandible, and occipital partitions contribute most significantly to skull sex identification.

3.2 Sex Identification Results for Incomplete Skulls

In practical situations, skulls exhibit various patterns of damage. Assuming a damaged skull contains only the mandible and frontal bone, this paper establishes a Logistic regression classification model incorporating both partitions:

$$P(y=1|x) = \frac{\exp(0.189 + 0.251X_1 + 0.179X_2)}{1 + \exp(0.189 + 0.251X_1 + 0.179X_2)}$$

Based on the established classification model, we performed leave-one-out cross-validation on skull samples and obtained comparative results for single-partition and combined-partition discrimination accuracy, as shown in .

From , we can see that the combined classification accuracy for mandible + frontal bone is higher than that of either partition alone. This demonstrates the applicability of the partition-based Logistic regression model under incomplete skull conditions. In practical applications, researchers can select appropriate partitions based on specific damage conditions to establish corresponding sex discrimination models for damaged skulls.

3.3 Comparative Analysis of Different Sex Identification Methods

The proposed method targets sex identification of incomplete skulls, a topic rarely addressed in literature. We selected methods from references [14] and [15] for comparison, evaluating their accuracy on both damaged and complete skulls. Results for partition numbers of 2, 3, 4, and 6 are shown in .

From , we observe that while methods in [14,15] achieve high accuracy for complete skulls, their precision decreases significantly when skulls are incomplete. With 2 partitions, their accuracies are only 73.3% and 74.7% respectively; with 4 partitions, they reach 82.1% and 83.2% respectively. In contrast, our method achieves 93.1% accuracy, approaching that of complete skull determination. The method in [14] and ours both belong to measurement approaches, establishing discriminant functions based on sexually dimorphic indicators. The method in [15] is an automatic recognition approach that builds a statistical shape model from all dense point clouds of the skull and uses Fisher discriminant after dimensionality reduction. The method in [14] measures only 13 indicators, lacking measurements for some partitions (e.g., parietal and occipital bones), making it effective for complete skulls but poor for incomplete ones. The method in [15] utilizes all point cloud information, achieving good results for complete skulls but poor performance for incomplete ones due to high dependence on sample integrity. Our method employs numerous measurement indicators distributed across the entire skull with low sample dependence, enabling identification using single-region or multiple-region combinations for any degree of skull incompleteness. It achieves excellent results for both complete and incomplete skulls, demonstrating greater reference and practical value.

---

4 Conclusion

Sex identification is the primary task in identity authentication. In actual criminal cases or excavated remains, incomplete skull fragments are frequently encountered. Sex determination of incomplete skulls can narrow investigation scope by 50% and provides guidance for ancient facial reconstruction. Traditional morphological methods largely depend on anthropologists' subjective understanding of population sexual dimorphism, with significant inter-observer variation, particularly among inexperienced observers. Measurement methods are more objective, convenient, and operational, making them the most commonly used approach. Therefore, this paper also employs measurement methods for incomplete skull sex identification.

We calculated measurable features, quantified non-measurable features, combined both feature types, and established a Logistic regression model to predict incomplete skull sex. Experiments demonstrate the proposed model' s utility for incomplete skull sex identification. The average classification accuracy across seven partitions is 75.6%. Combining the best-performing single regions (frontal bone and mandible) yields an average accuracy of 82.5%. Combining the four best-performing single regions achieves an impressive 93.1% average accuracy, approaching that of complete skull determination. Compared to other methods, our approach achieves good recognition rates for both complete and incomplete skulls, with particularly superior performance on incomplete skulls. Moreover, our measurement indicators are reasonably defined, distributed across the entire skull, and convenient to measure, providing high practical value.

However, this paper only quantified brow ridge prominence and external occipital protuberance prominence. Additional sexually dimorphic non-measurable features require quantification, such as frontal bone and nasal root morphology. Future work should quantify as many non-measurable features as possible and expand the skull sample size to further improve sex identification accuracy.

References

[1] Twisha S, Patel M N, Nath S, et al. Determination of sex using cephalo-facial dimensions by discriminant function and logistic regression equations [J]. Egyptian Journal of Forensic Sciences, 2016, 6(2): 114-119.
[2] Amores-Ampuero A. Sexual dimorphism in base of skull [J]. Anthropologischer Anzeiger, 2017, 74(1): 9-14.
[3] Ajanovic Z, Sarac-Hadzihalilovic A, Gojak R. Determination of sex discriminant function analysis of linear diameters in bosnian human skull [C]// Proc of CMBEBIH. 2017: 88-94.
[4] Shui Wuyang, Yin Rongchao, Zhou Mingquan, et al. Sex determination from digital skull model for the Han people in China [J]. Chinese Journal of Forensic Medicine, 2013, 28(6): 461-468.
[5] Li Chunbiao, Sun Eryu. A study on sex difference of adult skull of the northeast China by Fourier transform [J]. Journal of Anthropology, 1992, 11(4): 312-318.
[6] Li Ming, Fan Yinnan, Yu Yongmin, et al. Sex assessment of adult from southwest area of China by bones of facial cranium [J]. Chinese Journal of Forensic Medicine, 2012, 27(2): 132-134.
[7] Song Hongwei, Lin Ziqing, Sun Yan, et al. Study of sexual difference of single bone by discrimination function analysis [J]. Chinese Journal of Forensic Medicine, 1992, 7(1): 19-21.
[8] Ren Rongrong. Research on the method of sex identification based on 3D skull model [D]. Xi' an: Northwest University, 2017.
[9] Zhou Mingquan, Geng Guohua, Wu Zhongke, et al. Craniofacial morphology informatics [M]. Beijing: Science Publishing Company, 2015: 109-113.
[10] Xi Huanjiu. Anthropometric methods [M]. Beijing: Science Publishing Company, 2010: 51-78.
[11] Zhang Jizong, Shu Yongkang. Measurement methods of human skeleton [M]. Beijing: Science Publishing Company, 2007: 17-117.
[12] Zurifa A, Aida S H, Refet G. Determination of sex discriminant function analysis of linear diameters in bosnian human skulls [M]. [S.l.]: Springer Inter-

national Publishing, 2017: 88-94.
[13] Chovalopoulou M E, Bertsatos, Manolis S K. A comparative study based on the cranial sexual dimorphism of a modern greek population [M]. [S.l.]: Mediterranean Archaeology & Archaeometry, 2017, 17(171): 37-46.
[14] Liu Yuyong. The sex determination of han nationality in north China by adult facial skull X-ray [J]. Chinese Journal of Forensic Science, 2016, 84(1): 26-31.
[15] Luo Li, Wang Mengyang, Tian Yun, et al. Automatic sex determination of skulls based on a statistical shape model [J]. Computational and Mathematical Methods in Medicine, 2013, 251628: 1-251628: 6.

Note: Figure translations are in progress. See original paper for figures.