



ЦИФРОВАЯ АРХЕОЛОГИЯ DIGITAL ARCHAEOLOGY

Опыт подготовки материалов для систематизации и кластеризации наскальных изображений северных оленей (петроглифов) в долине реки Пегтымель, Чукотка

Experience in data preparation for systemization and cluster analysis of rock carvings of reindeer (petroglyphs) in the valley of the Pegtymel River, Chukotka Autonomous Okrug

Глотова Анастасия Павловна

Студентка магистерской программы «Цифровые методы в гуманитарных науках» НИУ ВШЭ
E-mail: anastasia.pavlovna.g@gmail.com

Anastasia P. Glotova

Student of MP «Digital Humanities»
HSE University

Локонцева Ксения Георгиевна

Студентка магистерской программы «Цифровые методы в гуманитарных науках» НИУ ВШЭ
E-mail: kseniya.lokonceva@mail.ru

Kseniya G. Lokonceva

Student of MP «Digital Humanities»
HSE University

Магомедов Ахмед Гусенович

Студент магистерской программы «Цифровые методы в гуманитарных науках» НИУ ВШЭ
E-mail: akhmed1@yandex.ru
ORCID: 0000-0002-6990-9859

Akhmed G. Magomedov

Student of MP «Digital Humanities»
HSE University

Сысоева Мария Антоновна

Студентка магистерской программы «Цифровые методы в гуманитарных науках» НИУ ВШЭ
E-mail: marusya.sysoeva@gmail.com

Maria A. Sysoeva

Student of MP «Digital Humanities»
HSE University

Резюме. В данной статье описываются первые результаты цифрового проекта, посвященного сохранению и систематизации в виде набора данных разрушающихся наскальных изображений северных оленей в долине реки Пегтымель для последующего анализа. Были осуществлены сбор, обработка, каталогизация и определены методологические аспекты кластеризации признаков для классификации изображений северных оленей. Последующий анализ предполагает решение научной проблемы, связанной с выявлением культурных кластеров пегтымельских петроглифов на основе стиля изображения оленей: по датам создания, принадлежности к культуре и народу, дифференциации индивидуального стиля и единой типологии.

Ключевые слова: цифровые гуманитарные науки, петроглифы, кластеризация признаков, геометрическая морфометрия, компьютерное зрение, GAC, нейросети

Abstract. The article describes the first results of a digital project dedicated to the preservation and systemization in the form of a dataset of crumbling reindeer rock carvings in the valley of the Pegtymel River for subsequent data development by other researchers. The collection, processing and cataloging of reindeer images were carried out, methodological aspects of clustering of features for the classification of reindeer were determined. The authors also attempt to solve the scientific problem of cultural clusters identification of Pegtymel petroglyphs based on the style of depiction of reindeer: by dates of creation, belonging to culture and people, differentiation of individual style and a single typology.

Keywords: digital humanities, petroglyphs, feature clustering, geometric morphometry, computer vision, GAC, neural networks

Предисловие

Авторы—студенты первого курса магистерской программы «Цифровые методы в гуманитарных науках» Национального исследовательского университета «Высшая школа экономики». В рамках обучения студенты применяют вновь полученные знания и навыки в реализации цифрового научного проекта.

Наша команда под руководством куратора, П.И. Лебедева, обратилась в Институт археологии РАН (г. Москва), где нам было предложено с помощью современных цифровых методов обработать, систематизировать и проанализировать наскальные изображения северных оленей из долины реки Пегтымель (Чукотка).

Выражаем огромную благодарность за научную консультацию и предоставление материалов Елене Сергеевне Левановой (канд. ист. наук, заведующая центром палеоискусства ИА РАН) и Юрию Михайловичу Свойскому (руководитель лаборатории RSSDA).

Данные и история изучения пегтымельских петроглифов

Пегтымельские петроглифы в Западной Чукотке—самое северное в Азии местонахождение наскальных изображений. Памятник на правом берегу р. Пегтымель, в устье ручья Кээнейкууль, был открыт в 1965 г. геологом Н.М. Саморуковым в районе горного массива Кэйныней. В дальнейшем пегтымельские петроглифы исследовались археологом Н.Н. Диковым, экспедициями Государственного Эрмитажа и ИИМК РАН.

В 2005–2008 гг. экспедиции под руководством д.и.н. Е.Г. Дэвлет, а затем в 2021 г. под руководством ее ученицы к.и.н. Е.С. Левановой провели комплексные работы по документированию памятника: расчистка поверхностей с петроглифами от лишайников, трасологические исследования, копирование поверхностей с рисунками, выявление разрушений и оценка антропогенного воздействия [1].

По результатам экспедиций были выделены 12 скоплений, содержащие 350 поверхностей с петроглифами. Большая часть скальных поверхностей с изображениями локализуется на протяжении 1500 метров. На наскальных рисунках изображены сюжеты повседневной жизни и верования древних северных народов, в том числе большое количество животных: северные олени, киты, медведи, собаки и др¹. Анализ этих изображений может позволить более точно связать изображения с какой-либо из археологических культур Севера.

В нашем распоряжении оказались прорисовки² всех поверхностей³, которые были сделаны по материалам экспедиций 2005–2008 гг. под руководством Е.Г. Дэвлет (авторы прорисовок—к.и.н. А.Н. Мухарева, Е.А. Миклашевич и А.М. Соротокина).



Илл. 1. Фото поверхности II-47 после расчистки. Сцена добычи оленей с каяков [2].

1 Подробнее об исследованиях и описании см. [2].

2 Первоначально делались копии на микалентную бумагу (метод протирки) и на прозрачную пленку, в дальнейшем пленки сканировались и прорисовки обрабатывались в Photoshop. В 2008 г. часть прорисовок делалась по цифровым фотографиям.

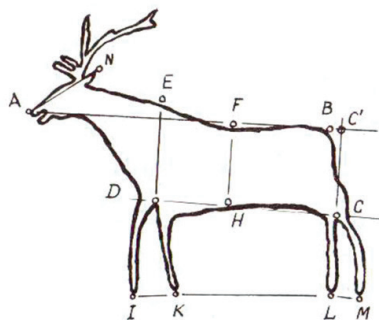
3 Каждая поверхность представлена отдельным файлом формата TIFF. В некоторых случаях особо крупные поверхности с большим количеством фигур разбивались на несколько файлов с нумерацией следующего уровня: а, б, с и т.д.



Илл. 2. Прорисовка поверхности ИI-47 [2].

Первым, кто поднял вопрос сохранения и обработки петгымельских петроглифов, был Н.Н. Диков. В своей книге 1971 г. он опубликовал прорисовки «от руки» семи поверхностей с изображениями из скопления I (см. прил. Б-1) и указал схему их расположения [3]. Также фотокопированием и прорисовками позднее занимались: Георгиевский И.Ю., Гиля Е.Ю., Дэвлет Е.Г., Миклашевич Е.А., Мухарева А.Н., Соротокина А.М.

В 1980 г. Яков Абрамович Шер в своей книге [4, с. 50], отмечая визуальные сходства и различия в признаках изображений животных, предложил формализовать язык описания изображений и перейти от субъективной оценки к математической модели, реализованной на ЭВМ. Первым шагом к строгой формулировке задачи классификации рисунков, по мнению Я.А. Шера, является эксплицитное изложение перечня признаков, по которому рисунки будут сравниваться между собой. Для этого было предложено на каждом изображении животного отмечать фиксированные базовые точки (см. Илл. 3). На этих точках могут быть построены вспомогательные линии, что позволяет сравнивать между собой изображения по единому перечню качественных и количественных признаков. Так, например, можно формализовать описание корпуса оленя, сравнивая длины перпендикулярных линий. Если $ED \sim FH \sim BC$, то это массивный корпус. Если $ED > FH > BC$, то мы имеем дело с поджарым корпусом. А если $ED > FH < BC$, то у корпуса имеется выраженная талия. Я.А. Шер приводит в общей сложности 25 признаков [4, с. 52], учитывающие как физиологические особенности оленя (характеристики корпуса, длина хвоста, длина головы и др.), так и особенности изображения (например, в какую сторону обращена фигура). Признакам присваиваются бинарные значения (1–наличие признака, 0–отсутствие признака).



Илл. 3. Базовые точки для описания контуров фигур оленей, предложенные Я.А. Шером [4, с.51].

Затем изображения сравниваются между собой по матрице показателей сходств (см. Таб. 1). Данная таблица иллюстрирует методику, предлагаемую Я.А. Шером для анализа изображений животных из разных локаций.

	1	2	3	4	5	6	7
1							
2	0,44						
3	0,41	0,80					
4	0,65	0,50	0,40				
5	0,04	0,08	0,10	0,20			
6	0,04	0,08	0,10	0,20	1		
7	0,65	0,28	0,35	0,79	0,04	0,04	

Таб. 1. Матрица показателей сходств. Частично воспроизведенная таблица из [4, с. 54]. Числа в первом столбце и первой строчке – порядковые номера рисунков. Числа на пересечениях столбцов и строк – величина показателя сходства между данной парой рисунков.

Цели и ожидаемые результаты проекта

Методика Я.А. Шера так и осталась по большей части теоретической. В этой связи на основе имеющейся методологии проект предлагает и применяет новые подходы к систематизации и кластеризации с использованием современных компьютерных технологий. Проект призван попытаться выявить культурные кластеры пегтымельских петроглифов на основе стиля изображения северных оленей. Предполагается, что изображения оленей являются наиболее информативными для первичного анализа пегтымельских петроглифов, так как северный олень – наиболее часто встречаемая фигура в пегтымельских петроглифах. Оленей больше, чем других фигур (антропоморфных и зооморфных), а стиль их изображения сильно варьируется. Из 350 поверхностей на 240 (69%) изображены олени. Общее количество идентифицированных по форме тела и наличию рогов изображений оленей – 723 (см. Илл. 4 и 5). В этой связи авторами предпринимается попытка сбора и анализа изображений северных оленей со всех 12 скоплений.



Илл. 4. Количество поверхностей с изображений оленей относительно общего числа поверхностей в скоплениях



Илл. 5. Количество фигур оленей в скоплениях

Ожидаемые результаты проекта:

- 1) качественный набор данных (т.н. датасет), включающий обрезанные изображения оленей в разных форматах, набор координат, таблицу с признаками;
- 2) кластеризация пегтымельских оленей по форме контуров;
- 3) сравнение трёх выбранных нами методов получения признаков для кластеризации⁴:
 - а) на основе ручного сбора и приведения признаков в табличный формат;
 - б) набора координат, описывающих форму цельных изображений оленей и только их брюшек, полученные методом геометрической морфометрии;
 - в) латентного пространства признаков оленей и ихбрюшек, полученного с помощью обучения генеративно-сопоставительной сети.

В данной статье описываются результаты первого этапа проекта. Полные результаты по завершению проекта также будут опубликованы в следующем году.

1. Подготовка датасета

Подготовка набора данных для анализа включала в себя две составляющие: обработка непосредственно самих изображений оленей и заполнение таблицы соответствующих им признаков.

1.1. Таблица признаков

С опорой на признаки Я.А. Шера были выявлены три категории признаков: каталогизирующие, логические и геометрические.

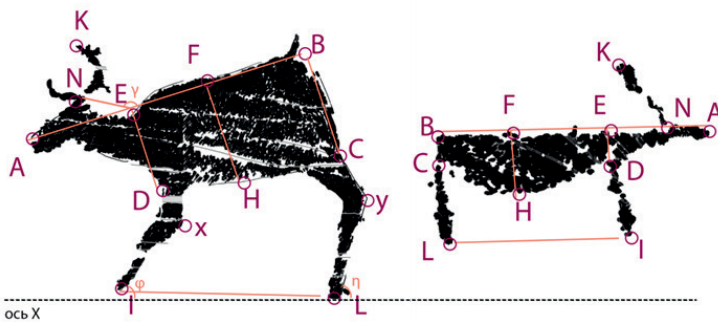
К каталогизирующим признакам относятся данные о расположении каждого оленя: скопление, поверхность и порядковый номер самого оленя.

К логическим признакам относятся такие бинарные атрибутивные показатели как целостность изображения, ориентация, расположение, поза (лежащий или стоящий). Также фиксируется наличие ушей, рогов, бородки, ног, хвоста. Рога могут иметь следующие признаки: утолщения, ответвления, направление по горизонтали или вертикали, прямая или полукруглая форма. Ноги характеризуются количеством, наличием изгиба в колене передней и задней ноги, типом копыт. Копыта представлены четырьмя вариантами: «пальцы», «пальцы с явно выраженными поноготками» и собранные воедино в форме «ботинок», «палок» и «треугольников» (см. прил. Б-2). В случае пальцев добавляется признак их количества. Есть среди них и признаки, относящиеся ко всему скоплению и указывающие на

⁴ Данные три метода были выбраны в качестве экспериментальных. Авторы открыты к предложениям по поводу применения других методов.

контекст: олень запряжен или проткнут палкой, изображен в многофигурной композиции, изображен среди прочих персонажей.

Геометрические признаки—это признаки реляционного характера: признаки, связанные с пропорциями животных. Для их получения на фигурах были выделены ключевые точки по Шеру (Илл. 6), линии между которыми позволяют представить оленей как отношения различных частей друг к другу. Кластерный анализ будет осуществляться в первую очередь машинными средствами: алгоритмами снижения размерности и проекции данных. Первоначально сформулированные отношения носят приблизительный характер и будут корректироваться в процессе анализа, но уже сейчас отражают основные группы, которые мы хотим выделить. Отношения можно поделить на три блока: туловище, голова и рога, ноги и копыта.



Илл. 6. Пегтымельские олени с разметкой ключевых точек для выявления геометрических признаков.

1.2. Признаки, относящиеся к туловищу:

Форма тела:

$BC/BE < 0,5$ —туловище узкое (т.е. ширина задней части туловища не превышает половину длины туловища, в противном случае туловище широкое);

$DE/BC < 0,3$ —туловище трапециевидное (когда ширина передней части туловища не превышает 0,3 от задней);

$DE/BE < 0,2$ —туловище треугольное.

Наклон тела:

α = угол BE к оси X (atan2)⁵

$\alpha > 3^\circ$ —тело развёрнуто вверх;

$-3^\circ < \alpha < 3^\circ$ —тело расположено горизонтально;

$\alpha < -3^\circ$ —тело наклонено вниз.

Выраженность брюшка:

$\beta = 2 * FH / (BC + ED)$

$\beta > 1,05$ —брюшко выпуклое;

$0,95 < \beta < 1,05$ —брюшко ровное;

$\beta < 0,95$ —брюшко вогнутое.

⁵ Рассчитывается арктангенс, т.е. угол между осью x и линией, проведенной из начала координат (0, 0) в точку с координатами (x, y). Угол определяется в радианах в диапазоне от -пи до пи, исключая -пи.

1.3. Признаки, относящиеся к голове и рогам:

Наклон головы:

β = угол AN к оси X (atan2)

$\beta >= 3^\circ$ —голова смотрит вверх;

$-3^\circ < \beta < 3^\circ$ —голова расположена горизонтально;

$\beta <= -3^\circ$ —голова наклонена вниз.

Длина головы:

$AN > AE/2$ —голова удлинённая;

$AN < AE/2$ —голова укороченная.

Выраженность головы:

(•)N \in AE—голова является продолжением тела;

(•)N \notin AE—голова выделена;

Наклон шеи относительно тела:

γ = угол NEB, обращённый к точке K

$\gamma = 180^\circ$ —шея параллельна туловищу;

$\gamma > 180^\circ$ —шея наклонена вниз;

$\gamma < 180^\circ$ —шея наклонена вверх.

Длина рогов:

$NK / AN \approx 1$ —рога выражены неявно;

$NK / AN > 1$ —рога длиннее головы.

1.4. Признаки, относящиеся к ногам:

Отношение длин передних и задних ног⁶:

$DI/CL \approx 0$ — ноги приблизительно равны;

$DI/CL > 0$ —передняя нога длиннее;

$DI/CL < 0$ —задняя нога длиннее.

Расположение ног относительно друг друга:

ϕ = угол DI к оси X

η = угол CL к оси X

$\phi \approx \eta$ — ноги параллельны;

$\phi \neq \eta$ — ноги расходятся.

Все эти признаки могут быть связаны с контекстом создания изображений, уточнять род животного из семейства оленьих, указывать на принадлежность к домашним или диким⁷. Наличие рогов и объёмы туловища могут свидетельствовать о сезоне. Так, например, взрослые самцы остаются без рогов зимой, а молодняк—весной. Безрогие самки северных оленей—явление встречающееся, но не частое. Каждый тип копыт также скорее всего указывает на определённый сезон, например, зимой для препятствия скольжению оленьи копыта покрываются плотным слоем волос, что может быть иллюстрировано копытами типа «ботинки» и «треугольники». Наличие бородки и массивной морды поможет выявить лосей.

1.5. Получение размеров

Разметка осуществлялась с помощью написанной куратором Павлом Лебедевым программы⁸ (см. прил. Б-3, Б-4). Заданный набор точек можно использовать для других алгоритмов и пополнять список соотношений в ходе анализа.

⁶ Если присутствует пара передних или задних ног, отрезки DI и CL откладываются посередине между ними. Для суставов на ногах предложены точки x, y, но в текущих признаках они не используются, и есть не у всех оленей.

⁷ В чукотском языке это различие находят выражение в использовании двух разных терминов: борапы, означающего домашнего оленя, и ыльвлю, обозначающего дикого северного оленя [5].

⁸ Код, написанный на языке программирования Python, позволяет производить разметку всех загруженных изображений в ручном режиме. Доступные опции: подпись курсора, навигация по принципу галереи, отмена мет-

Из полученной таблицы координат интересующих нас точек были получены расстояния между ними. Из-за того, что наши прорисовки имеют разный масштаб, невозможно взять координаты полученные в ходе разметки, вычислить расстояния и сравнить их. Чтобы измеренные расстояния было можно сравнивать, была выполнена операция приведения размеров для каждого оленя. За единицу был взят размер от кончика носа до хвоста (AB). Операции по вычислению и приведению размеров были выполнены с помощью языка Python⁹.

1.6. Предобработка изображений

Первичная обработка изображений входила в подготовку изображений для каждого варианта анализа. Она включала в себя:

1. Вырезание оленей из скоплений;
2. Фильтрация изображений;
3. Приведение к квадратному формату;
4. Корректировка размера изображений.

Олени вырезались вручную на самом первом этапе при использовании таких графических редакторов, как Paint и AdobePhotoshop.

Следующие этапы проводились с помощью языка программирования Python и библиотеки для работы с изображениями Open CV. Изображения были отфильтрованы от не относящейся к прорисованным чёрным оленям информации: это прорисованные серым утраты (природные трещины и сколы) скальных поверхностей. Для этого выделенный диапазон серых пикселей заменялся на белые. Далее олени обрезались по координатам чёрных пикселей и, исходя из горизонтальной или вертикальной ориентации оленя, изображения дотраивались до квадратного формата. После этой обработки изображения сохранялись в формате png в размере, необходимом для каждого типа задач¹⁰.

1.7. Выявление контуров и получение однородного изображения

В данной части будут описаны методы компьютерного зрения для обработки прорисовок петроглифов. В связи с тем, что мы работаем с прорисовками изображений, а они являются неоднородными и зачастую представляют из себя рисунки из не связанных между собой мазков, перед нами встала задача выделения контуров. Это было необходимо для проведения эксперимента с методами геометрической морфометрии [7].

Для автоматизации процесса выделения контуров был выбран язык программирования Python. Первоочередными задачами стали: выбор библиотеки для работы с изображениями и метода выделения контуров. В результате были отобраны девять библиотек (NumPy, SciPy, scikit-image, Pillow, Open CV, SimpleCV, Mahotas, Simple ITK, pgmagick). После изучения документации и функционала, связанного с выделением контуров, было принято решение писать код с помощью scikit-image. Данная библиотека выбрана по ряду причин: работает с массивами NumPy, реализовано много сложных математических алгоритмов для работы с изображениями, хорошо задокументирована, имеет много практических примеров¹¹. Решающим фактором в выборе стало то, что изображения scikit-image представлены в виде массивов (ndarrays) NumPy, следовательно, многие общие операции могут быть выполнены с использованием стандартных методов NumPy для работы с массивами, которые хорошо знакомы нашей команде.

ки, выбор метки. Координаты каждого изображения автоматически сохраняются в csv-файл. <https://github.com/Petroglyphs/petroglyphs/blob/main/markup.py>

9 URL: <https://github.com/Petroglyphs/petroglyphs/blob/main/sizes.ipynb>

10 URL: https://github.com/Petroglyphs/petroglyphs/blob/main/Image_adjustment.ipynb

11 URL: <https://scikit-image.org/>

Изначально полагалось, что достаточно будет сегментировать изображения с последующим размытием, однако данный подход привел бы к утрате исходной формы прорисовок, это было недопустимо. Изучение математических алгоритмов геометрической морфометрии помогло лучше понять, что именно необходимо искать.

Существует множество методов выделения контуров: самый известный, пожалуй, это алгоритм обнаружения Canny¹², Робертса, Превитта и Собеля¹³, морфологические змеи (Morphological Snakes)¹⁴, активные контурные модели¹⁵, также называемые змеями.

Первым из потенциально подходящих был детектор границ Canny, однако, при более детальном рассмотрении, выяснилось, что алгоритм подходит только для хорошо прорисованных оленей без разрывов, большой проблемой являлось выделение любых границ внутри контура. В результате Canny был исключен до проведения эксперимента.

Второй на очереди была модель активного контура. Змеи представляют собой основу в компьютерном зрении, были представлены Майклом Кассом, Эндрю Уиткиным и Деметри Терзопулосом для выделения контура объекта из возможно шумного 2D изображения. Модель змей популярна в компьютерном зрении, змеи широко используются в таких приложениях, как отслеживание объектов, распознавание форм, сегментация, обнаружение краев и стереосопоставление. Змея—это минимизирующий энергию, деформируемый сплайн¹⁶. Если описывать механизм работы змей простым языком, то это подвижный сплайн (линия) который притягивается к контурам объекта в соответствии с выставленными ограничениями и одновременно сопротивляется деформации.

Змеи не решают всей проблемы нахождения контуров на изображениях, так как метод требует предварительного знания желаемой формы контура. Скорее, они зависят от других механизмов, таких как взаимодействие с пользователем, взаимодействие с некоторым процессом понимания изображения более высокого уровня или информации из изображения, смежных во времени или пространстве.

Для выбора метода было взято одно предварительно вырезанное изображение оленя без серого цвета в исходном размере примерно 3000×2500 pix.

Эксперимент с активной контурной моделью(змеями) тоже закончился неудачей, кроме того, код работал очень медленно. Это, скорее всего, связано с тем, что метод змей по умолчанию имеет различные ограничения и угловые случаи, когда сходимость выполняется плохо. Существует несколько альтернатив, которые решают проблемы метода по умолчанию, хотя и со своими нюансами.

Одной из частных альтернатив являются морфологические змеи, и при отборе методов ставка делалась именно на них. Если не вдаваться в подробности, это семейство методов сегментации изображений¹⁷, их поведение похоже на поведение активных контуров, но они используют морфологические операторы¹⁸ (такие как расширение или эрозия) над двоичным массивом вместо решения дифференциальных уравнений в частных производных над массивом с плавающей запятой, что является стандартным подходом для активных контуров.

12 URL: <https://russianblogs.com/article/80931038618/>

13 URL: <https://habr.com/ru/post/114452/>

14 URL: https://scikit-image.org/docs/dev/auto_examples/segmentation/plot_morphsnakes.html#sphx-glr-auto-examples-segmentation-plot-morphsnakes-py

15 URL: https://ru.wikibrief.org/wiki/Active_contour_model

16 Подробно о минимизации энергии и механизме работы модели см.: URL:<https://russianblogs.com/article/21522271457/>

17 URL: https://neerc.ifmo.ru/wiki/index.php?title=Сегментация_изображений

18 URL: <https://habr.com/ru/post/113626/>; https://intuit.ru/studies/courses/993/163/lecture/4505?page=6#:~:text=Морфологические_операторы

Это делает морфологических змей более быстрыми и численно более стабильными, чем их аналоги.

Семейство методов морфологических змей имеет несколько реализаций, две из которых оказались доступны в `scikit-image`:

1) Активные контуры без краёв или модель диффузной змеи—Active Contours without Edges (далее ACWE)¹⁹. Реализованы в функции `morphological_chan_vese`—не требует четкого определения контуров объекта, работает с исходным изображением без предварительной обработки. Требуется, чтобы внутренняя часть объекта в среднем выглядела иначе, чем внешняя (т. е. внутренняя область объекта должна быть в среднем темнее или светлее, чем внешняя область)

2) Геодезические активные контуры или геометрические активные контуры или конформные активные контуры—Geodesic Active Contours (далее—GAC), они и были опробованы²⁰.

По описанию нам больше подходил метод GAC, но поставленные задачи включали в себя и изучение всех представленных вариантов обработки.

Модель ACWE основана на технике эволюции кривых, функционале Мамфорда—Шаха²¹ для сегментации и набора уровней²². Модель может обнаруживать объекты, границы которых не обязательно определяются градиентом. Контур выделяется, опираясь на теорию множества уровней²³, но вместо эволюции²⁴ активного контура мы имеем поток средней кривизны²⁵, который останавливается на желаемой границе, однако условие остановки не зависит от градиента, как в классических моделях активного контура, а связано с определенной сегментацией изображения.

В ACWE вы можете задать: количество итераций—`iterations`, набор уровней—`init_level_set` (тут есть только 2 варианта шахматная доска и круг), сглаживание—`smoothing`, вес внешней области—`lambda1`, вес внутренней области—`lambda2`, и функция для отладки—`iter_callback` (см. в прил. А), `morphological_chan_vese(image, iterations, init_level_set='checkerboard', smoothing=1, lambda1=1, lambda2=1, iter_callback=<function <lambda>>)`.

Результаты пробной сегментации:

Morphological ACWE segmentation



Morphological ACWE evolution



Илл. 7.

19 URL: <https://www.math.ucla.edu/~lvese/PAPERS/IEEEIP2001.pdf>

20 О реализации ACWE и GAC см.: URL: https://scikit-image.org/docs/dev/auto_examples/segmentation/plot_morphsnakes.html#sphx-gl-auto-examples-segmentation-plot-morphsnakes-py

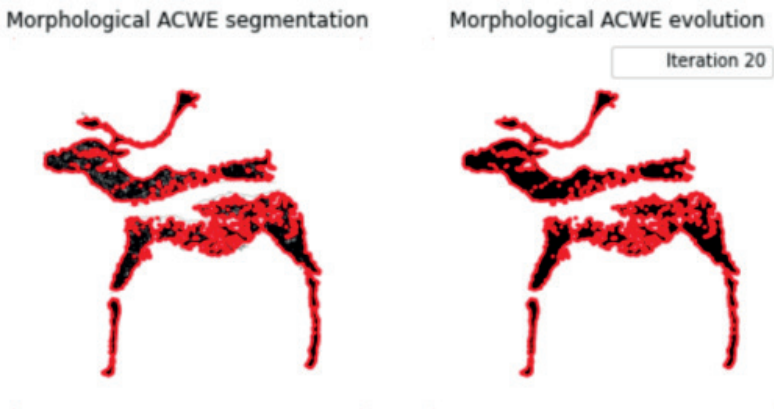
21 URL: https://en.wikipedia.org/wiki/Mumford%E2%80%93Shah_functional

22 Levelset – математическое обоснование: URL: https://en.wikipedia.org/wiki/Level_set

23 URL: https://lyceum.urfu.ru/fileadmin/user_upload/AnufrienkoComb.pdf; https://ru.wikipedia.org/wiki/Множество_уровня

24 URL: <https://russianblogs.com/article/21522271457/>

25 URL: https://ru.wikipedia.org/wiki/Поток_средней_кривизны



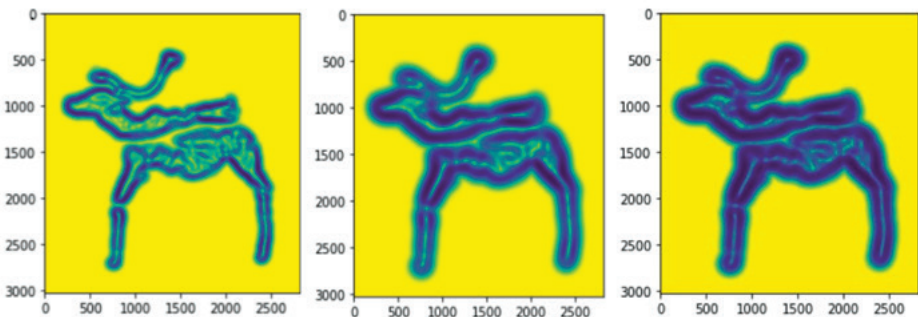
Илл. 8.

Метод отлично показал себя на части изображений (Илл. 7), но на прорисовках с большими разрывами не работал так как необходимо, предположительно это связано с тем, что метод основан на сегментации и никак не связан с градиентом (Илл. 8).

Второй метод **GAC** использует идеи из укорачивания евклидовой кривой эволюции. Поток сокращения кривой (Curve-shortening flow)—это процесс, который сжимает гладкую кривую на евклидовой плоскости на основе ее кривизны²⁶. Контурные разделяются и объединяются в зависимости от обнаружения объектов на изображении. Эти модели в значительной степени основаны на наборах уровней GAC (levelset, упомянутый выше в ACWE, и метод level set method²⁷) и активно используются в обработке медицинских изображений.

Реализован в функции `morphological_geodesic_active_contour`, подходит для изображений: с видимыми контурами, контуры могут быть шумными, загроможденными, частично нечеткими и прерывистыми границами. GAC более требователен, на вход принимает только предобработанные изображения, а именно, требуется инверсия величины градиента, и эффективность метода почти полностью зависит от этого шага (Илл. 9–11).

В нашем случае инверсия была произведена с помощью функции `inverse_gaussian_gradient(image, alpha=number, sigma=number)`



Илл. 9.alpha=150, sigma=30

Илл. 10. alpha=1000, sigma=50

Илл. 11.alpha=3000, sigma=45

²⁶ Поток сокращения кривой – теоретическое обоснование: URL: https://ru.wikibrief.org/wiki/Curve-shortening_flow

²⁷ URL: https://en.wikipedia.org/wiki/Level-set_method

Параметр α управляет крутизной инверсии. Большее значение сделает переход между плоскими областями и граничными областями более крутым в результирующем массиве. Чем больше α тем плавнее будет переход, чем меньше α тем больше шансов что будет разрыв между частями изображения. К выбору величины параметра стоит подходить аккуратно, слишком большая величина чрезмерно размывает изображение.

Параметр σ — стандартное отклонение фильтра Гаусса, примененного к изображению²⁸. При сложных формах изображения (согнутые конечности, рога вдоль морды) не стоит выставлять большое значение, будет утрачена детальность.

Обнаружить ошибку выбора параметров возможно только после получения контуров (Илл. 12–14).



Илл. 12. Чрезмерно размытое изображение

Илл. 13. Исходное изображение

Илл. 14. Идеальный вариант

После предобработки можно применять GAC.

В ходе эксперимента было выявлено, что достаточно оперировать четырьмя параметрами: количеством итераций (iterations); сглаживанием (smoothing)—количество применений оператора сглаживания за итерацию, большие значения приводят к более плавной сегментации, очень важно не забывать про прямо пропорциональную зависимость от количества итераций и сглаживание по Гауссу во время предобработки; порог (threshold)—области изображения со значением меньше этого порога будут считаться границами, выделение контура остановится в этих областях; сила притяжения «шар» (balloon)—направляет контур в неинформативные области изображения, т. е. в области, где градиент изображения слишком мал, чтобы подтолкнуть контур к границе, отрицательное значение уменьшит контур, а положительное значение расширит контур в этих областях. Установка этого значения на ноль отключит силу воздушного шара. Данный параметр не стоит увеличивать для сложных контуров и имеет смысл сделать больше на прорисовках с разрывами, однако можно добиться хороших результатов полностью за счет других параметров.

MorphGAC—полностью выполняет поставленную задачу выделения контуров при правильно заданных настройках и обдуманной предобработке. Сложность в выборе параметров заключается в том, что одного и того же результата можно добиться разными способами, поэтому важно менять параметры осознанно. При работе с GAC следует стараться добиться максимально хорошего результата за счет фильтра Гаусса, чтобы минимально изменять параметры в самом методе.

Ввиду разнообразия параметров перед обработкой всего массива изображений, чтобы подобрать наиболее подходящие параметры для всех изображений, были выбраны пять максимально разнообразных прорисовок, в которых содержались разные «проблемные» формы (Илл. 15).

²⁸ URL: https://ru.wikipedia.org/wiki/Размытие_по_Гауссу



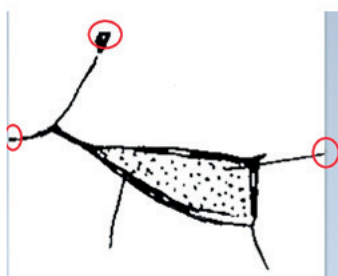
Илл. 15. Исходные изображения



Илл. 16. Результат отработки кода, взятый для обработки всей серии.

При запуске пробной серии был выявлен ряд проблем:

- серый цвет, отражающий фактуру поверхности, выделяется в контур (как мы и предполагали);
- слишком долгая отработка кода;
- контуры не выделяются или выделяются некорректно, если вырезка происходила вплотную по контуру оленя (Илл. 17, 18), необходимо оставлять не менее двух пустых пикселей от объекта, контур которого планируется получить.



Илл. 17



Илл. 18

Перед запуском финальной серии, состоящей из 723 вырезок, был применен код для очистки от серого и уменьшению размера изображения до 600×600 pix, написанный в Open CV. Библиотека Open CV отлично сочетается с scikit-image, важно помнить, что в Open CV используется кодировка цвета BGR и тип изображения uint8, в scikit-image кодировка цвета RGB а изображения—массивы NumPy, поддерживающие различные типы данных.

В процессе предобработки изображений из выборки были убраны пять изображений, не подходящих для GAC, код с подобранными на тестовой выборке параметрами был запущен на 719 вырезанных прорисовках оленей. Некоторые (сомнительные) прорисовки были запущены в нескольких итерациях, чтобы имелась возможность выбрать лучший результат (Илл. 19–22).

После первой итерации получены 275 выделенных непрерывных контуров, осталось 444.

В результате второй итерации—255 (16 изображений задублировано с первым результатом), осталось 189.

В результате третьей итерации получено 108 изображений, осталось 102; позже планируется найти подходящие параметры и для них (3 изображения задублировано с первым результатом).



Илл. 19. Исходное изображение
Илл. 20. После первой итерации
Илл. 21. После второй итерации
Илл. 22. После третьей итерации

Для максимального охвата подаваемых GAC на вход изображений в отдельные выборки следует выбирать изображения с большими разрывами, маленькими расстояниями между частями изображения, и тонкими линиями. При необходимости максимально сохранить исходный контур, как в нашем случае, следует идти от более жестких настроек к более мягким.

2. Методы получения признаков

2.1. Геометрическая морфометрия

Одной из задач была проверка возможностей метода геометрической морфометрии для поиска отличий формы.

Геометрическая морфометрия появилась в зоологии как развитие метода классической морфометрии. В отличие от классической морфометрии, опирающейся на линейные измерения дистанций между характерными точками.

Метод геометрической морфометрии позволяет численно описывать форму, используя координаты меток–ландмарков, которые расставляют на сравниваемых морфологических объектах. Различие между объектами по их форме определяется как разница между конфигурациями меток, где «форма является многомерным морфологическим признаком, вариации которой характеризуют состояние этого признака» [7, с.15].

Геометрическая морфометрия разрабатывается как особый аналитический инструмент, позволяющий исключить влияние размеров на результаты анализа формы морфологических структур. Ключевым понятием геометрической морфометрии является понятие о форме объекта, данное Дэвидом Кендаллом (Kendall, 1977). По Кендаллу форма – это вся геометрическая информация об объекте за исключением информации о его местоположении, ориентации в пространстве координат и размерах. Таким образом данный метод не учитывает поворота фигур, их размеров (увеличения, уменьшения) и смещения (расположения предметов на плоскости) [8].

Через снижение мерности данных, представляющих координаты точек, мы получаем уникальный центроид каждой формы и распределение данных центроидов в виде точек на поверхности «кэндаллового пространства» (пространство всех возможных форм в виде гиперсферы). Таким образом, проецируя полученный результат на тангенциальную плоскость, мы можем отследить, насколько схожи или отличны те или иные формы.

Технология описания формы применяет два типа точек, которые мы можем извлечь, – это метки (landmarks) и полуметки (semilandmarks). Метки расставляются, опираясь на гомологичные элементы формы, и описывают её отдельные структурные элементы, которые являются значимыми для нашего анализа. Набор полуметок «характеризует форму гладких, шарообразных или эллипсоидных объектов, а также тех, у которых имеется сложная структура, способная существенно изменяться и варьировать у разных особей» [7, с. 49]. Они располагаются независимо от особенностей формы на равном расстоянии друг от друга и описывают её целиком.

На первом этапе анализа должна быть исключена информация, не имеющая отношения к собственно форме. Для удаления избыточной информации применимы несколько способов, называемых методами суперимпозиции [7, с. 69]. Отличаясь по реализации, методы суперимпозиции приводят к совмещению исходных конфигураций в ходе аффинных преобразований – трансляции, вращения и масштабирования.

Подвергнутые суперимпозиции конфигурации меток (и полуметок) сравниваются относительно консенсуса – конфигурации меток, полученной в результате усреднения координат каждой метки.

Изначально планировалось описывать форму каждого из оленей целиком, т.е. с помощью полуметок, поэтому мы имели специфические требования к формату изображений.

Требования были продиктованы применением автоматизированной расстановки полуметок с помощью утилиты tpsDig (Rolf)²⁹. Обязательное условие корректной работы алгоритма – у изображений должны быть выделены замкнутые контуры, описывающие силуэт,

залитый одним цветом. В разделе «Предобработка изображений» описан процесс подготовки исходного формата файлов и фигур для метода выделения координат с помощью полуметок.

При апробации подхода были выявлены следующие ограничения:

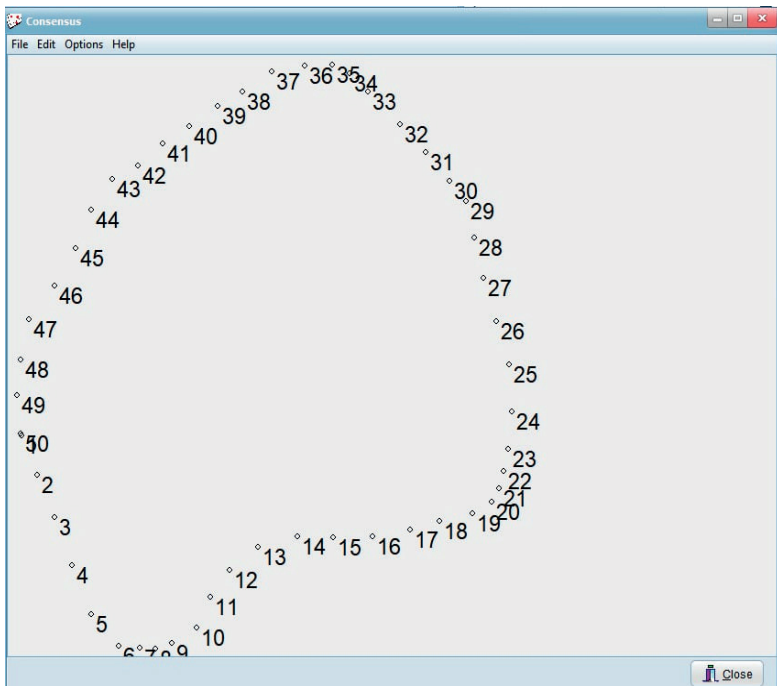
1) Из-за сложной, негомологичной формы оленей мы не можем добиться того, чтобы полуметки имели начало в одной и той же точке, но это принципиально важно для корректных результатов дальнейшего анализа. В дальнейшем для коррекции существующего алгоритма возможно написание собственного программного обеспечения.

2) Обязательное требование к формату изображений – чёрный залитый силуэт оленя на белом фоне – приводит к потере оригинальности прорисовки петроглифа. То есть мы преобразуем аутентичное изображение, которое является историческим источником и научной копией объекта культурного наследия. Однако наша гипотеза заключается в том, что эта процедура не внесёт глобальной погрешности в результаты кластеризации.

3) У данной программы имеется существенное ограничение в формате изображений – качественное, сильно тяжелое по объему изображение, которое при расстановке включает в себя более 6000 полуметок по контуру, вызывает ошибку. В ходе подготовки формата изображений, описанной в разделе выше, это ограничение было преодолено.

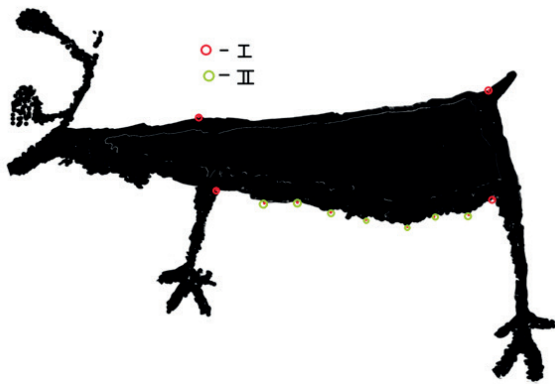
4) Ограничения в анализе количества полуметок – программное обеспечение составлено с учётом того, что принимает не более ста координатных точек. Если же полуметок больше ста штук, программа не сможет обработать данные. Опять же мы приходим к тому, что есть потребность написания программного обеспечения под данную узкую задачу.

На 50 полуметках (semilandmarks) анализ показал только максимально общие деформации фигур (Илл. 23), однако так как полуметки исходили из разных структурных частей фигур, мы не можем отследить, что именно менялось в данном случае:



Илл. 23. Консенсус формы оленей по контурам на 50 полуметках с использованием tpsRelw(w64)

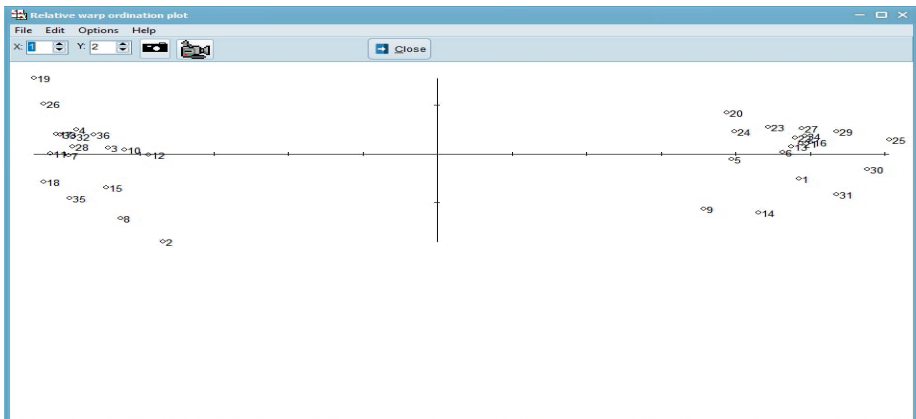
После чего алгоритм получения данных с координатами был пересмотрен. Во время проверки метода алгоритм получения данных и дальнейшие их преобразования проводились без отступления от традиционного использования в биологических науках. Путём проб и ошибок была выбрана наиболее гомологичная форма для всей выборки фигур оленей–туловище. Оно есть практически у всех фигур за редкими исключениями, мы можем выделить на нем единообразные структурные точки, оно является основой формы фигуры. В качестве набора координат были выбраны метки I и II типов (см. Илл. 24): метки I типа отражают места на сочленении структурообразующих элементов фигуры, если мы говорим об анатомическом понимании оленя. Их в наших фигурах было 4: под передней ногой, под задней ногой, на лопатке и на крупе. А метки II типа–описывают контуры наибольшей кривизны формы, в данном случае это контур брюха оленей, включающий всего 7 меток. Особенностью метода является то, что метки должны быть расставлены в одном порядке на каждой из фигур.



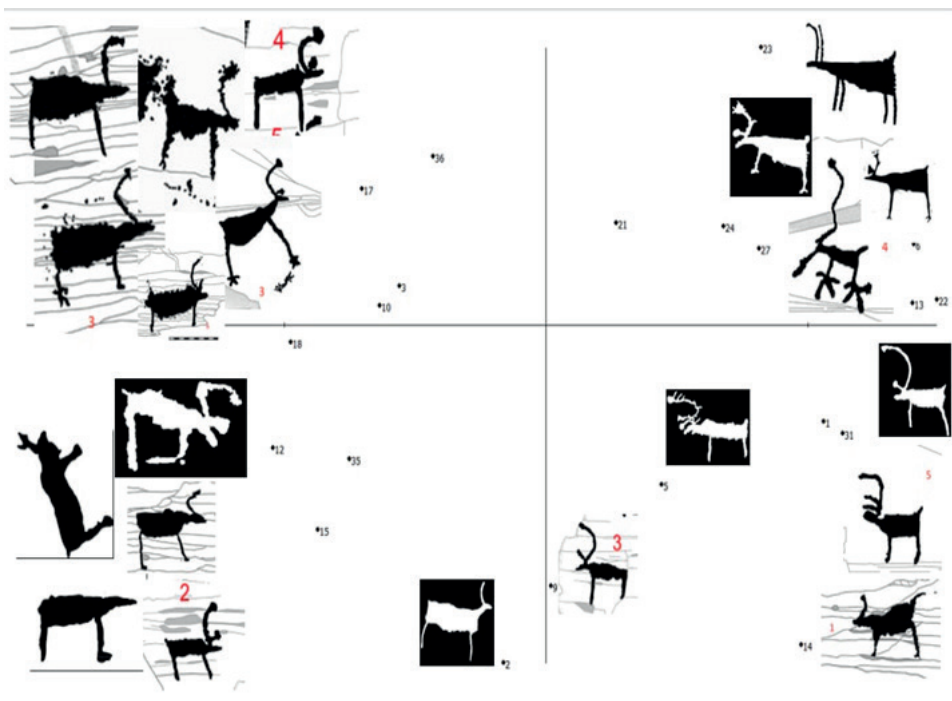
Илл. 24. Принцип разметки туловища оленя для метода геометрической морфометрии.

Первая оценка графика вариативности форм была проведена в программе tpsRelw на небольшой выборке в 30 фигур. Далее с помощью метода суперимпозиции методом Букштейновых координат формы [9] была снижена размерность данных и получены значения центроидов форм, которые мы смогли спроецировать на плоскость.

В данном случае график отразил две максимально не схожие между собой по форме группы (Илл. 25). Relative warp ordination plot–это многомерная ординация координат частных деформаций методом главных компонент (PCA).



Илл. 25. График частных деформаций (tpsRelw).



Илл. 26. Интерпретация результатов графика частных деформаций.

При интерпретации данных результатов выяснилось, что в ходе анализа кластеризация оленей происходит по повороту фигуры (Илл. 26). Однако, метод геометрической морфометрии не должен учитывать отражение и анализировать его. Возникли вопросы, а коррелирует ли ориентация фигуры с формой оленя? Является ли группа оленей, обращенных вправо, схожей между собой без учета ориентации, и отличной от оленей, обращенных влево?

Основная **гипотеза** была сформулирована следующим образом:

Корреляция формы и ориентации ложная и является артефактом применяемого метода суперимпозиции.

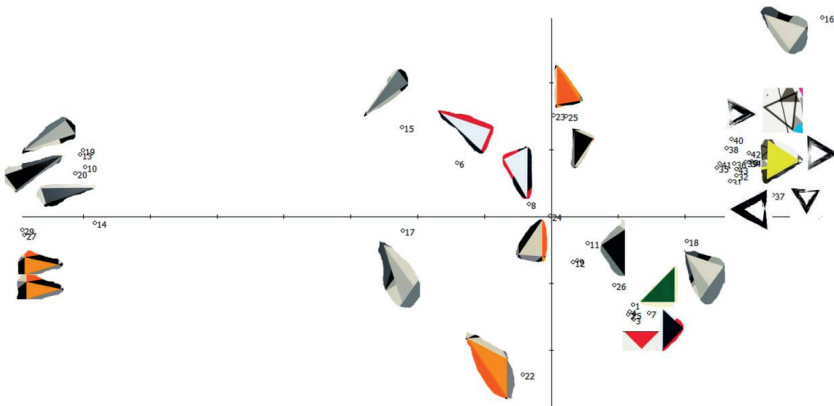
Контргипотеза:

Корреляция формы и ориентации истинная. В таком случае она должна сохраняться при применении иного метода суперимпозиции.

Для проверки гипотезы была составлена выборка из 43 треугольников разной формы и разных поворотов вершины, на которых также были расставлены метки, обозначающие координаты x и y . В данном случае метки 1 типа располагались на углах по порядку, начиная от вершины вправо, а метки 2 типа делили стороны треугольников пополам.

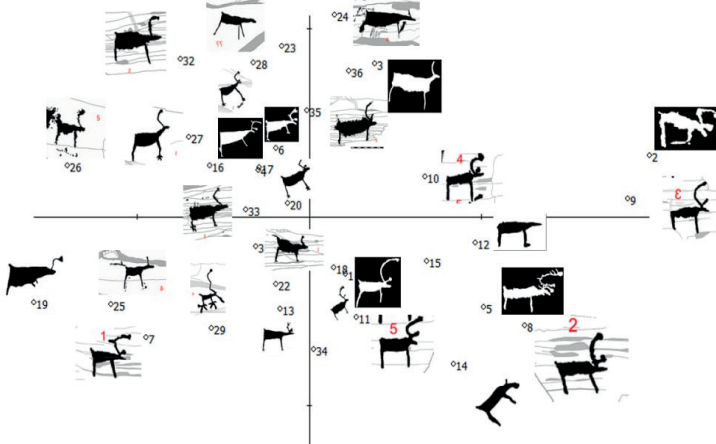
На Илл. 27 видно, что схожие по форме фигуры (например, равнобедренные треугольники), вне зависимости от того, в какую сторону они обращены, собираются вместе по признаку схожести их формы.

Таким образом, подтверждается утверждение о толерантности метода геометрической морфометрии к отражению фигур.



Илл. 27. График распределения частных деформаций треугольников разной формы и поворота.

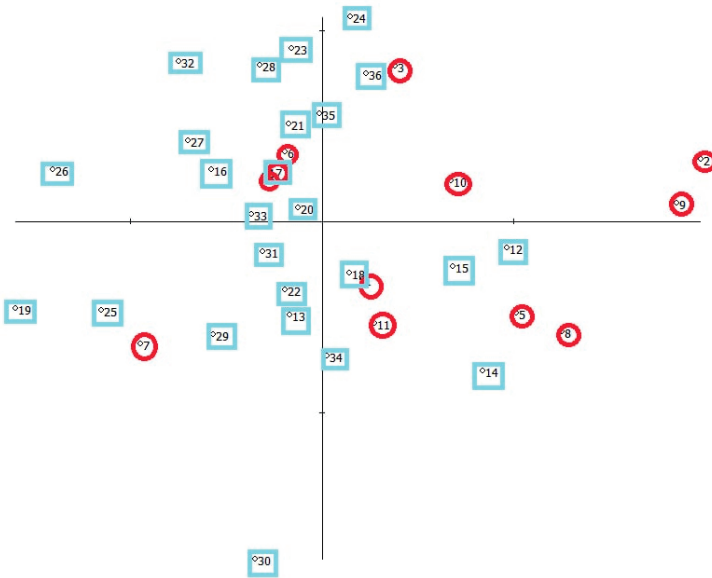
Однако применяемое программное обеспечение tpsRelw использует для анализа метод Букштейновых координат формы [9], при котором преобразования производятся относительно осевой линии, в качестве конечных точек которой произвольно выбираются две лантмарки, которым присваиваются координаты (0,0) и (0,1) соответственно. Все конфигурации в выборке совмещаются друг с другом таким образом, что координаты базовой линии лантмарок принимают значение (0,0) и (0,1) [7]. В данном случае точки осей были выбраны автоматически и таким образом, что поворот стал значим для разделения, зеркально отраженные фигуры распределились по двум противоположным кластерам.



Илл. 28. Распределение частных деформаций оленей при обращении всех фигур вправо.

В результате поворота всех оленей в одну сторону и проведения очередного анализа было обнаружено, что разделение на 2 явных группы исчезает и олени распределяются менее явными образом вокруг нулевой точки осей, сохраняя, однако, распределение по квадрантам в соответствии с преобладающей формой туловища (Илл. 28).

В дальнейшем также была проверена корреляция распределения форм оленей в соответствии со скоплениями. На графике может быть отмечена тенденция тяготения оленей из различных скоплений в противоположные квадранты, однако наблюдается существенное взаимное проникновение групп (Илл. 29).



Илл. 29. График частных деформаций произвольно выбранных оленей из I и II скопления. I скопление помечено красным кругом, а II скопление – голубым квадратом.

В результате проверки гипотез можно сделать следующие выводы.

График отражает, что поворот оленя (ориентация влево или вправо) прямо коррелирует с формой его туловища, однако метод суперимпозиции с применением Букштейновых координат при случайном выборе базовой линии ландмарков может приводить к возникновению ложных корреляций, увеличивающих или уменьшающих имеющиеся различия в форме. Поэтому результат может быть определен как артефакт метода суперимпозиции.

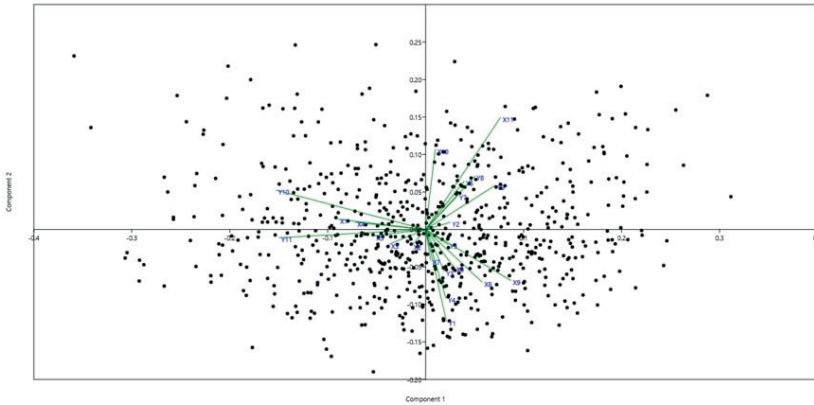
Графически наблюдается нечеткая корреляция между формой туловища и номером скопления.

Метод геометрической морфометрии проявляет толерантность к аффинным преобразованиям формы, как и было заявлено в источниках о нем.

Следующим этапом работы стала проверка метода на полной выборке. Общая выборка фигур на данный момент включила в себя 723 оленя, но путём визуального анализа для данного метода было выбрано 698 фигур, которые включают изображение туловища (в общей выборке есть часть фигур с изображением только головы и шеи, которые были исключены из неё).

Метки были расставлены по тому же принципу, как и в тестовой выборке (Илл. 24), однако снижение мерности полученных координат производилось с помощью альтернативного метода – Обобщенного Прокрустова анализа (GPA – Generalized Procrusted Analysis), в противоположность методу Букштейновых координат формы. Этот метод суперимпозиции в геометрической морфометрии считается одним из наиболее распространенных и опирается на другой принцип совмещения конфигураций. В данном случае конфигурации совмещаются путем трансляции их центроидов – точек координаты X и Y которых представляют собой усредненные значения всех X и Y координат каждой конфигурации соответственно в точку (0,0). Далее происходит масштабирование и вращение конфигураций до минимизации расстояния между гомологичными точками. Точкой сравнения является эталонная форма – консенсус всех фигур выборки (Consensus). График показывает вариации деформаций (численного выражения степени различия формы) относительно эталона и поворот фигуры

не влияет на распределение. График на Илл. 30 отражает вклад в деформацию отдельных координат, здесь представленных векторами, исходящими из центра пересечения осей. Чем длиннее вектор, тем более значимый вклад в деформацию форм туловищ оленей он вносит. Как мы видим, это точки 10 и 11, которые соответствуют лопатке и крупу оленя, а также чуть меньше выражены 9 и 1 (точки под задней и передней ногой соответственно). В данном случае все они являются метками 1 типа.

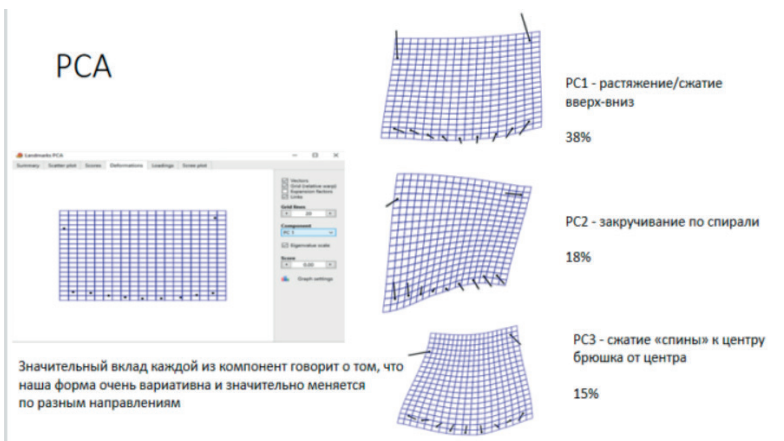


Илл. 30. График деформаций выборки в 698 оленей, полученный с помощью GPA и PCA.

На графике (Илл. 30) PCA не отражает явно выраженных групп, скорее он похож на распределение, полученное в тестовой выборке с помощью суперимпозиции методом Букштейновых координат формы (Илл. 28).

Для определения вклада каждой из компонент в деформацию также была проанализирована сетка деформаций по каждой из 3-х основных компонент в программе PAST (Paleontological statistics software package for education and data analysis)³⁰.

Вклад каждой из компонент оказался довольно значим (Илл. 31), поэтому мы можем сказать о том, что форма очень вариативна, а наша выборка не склоняется к деформациям по какой-то определенной компоненте. Скорее, в ней практически одинаково выражены различные типы деформаций, однако наибольший вес составляет растяжение и сжатие формы туловища по толщине.



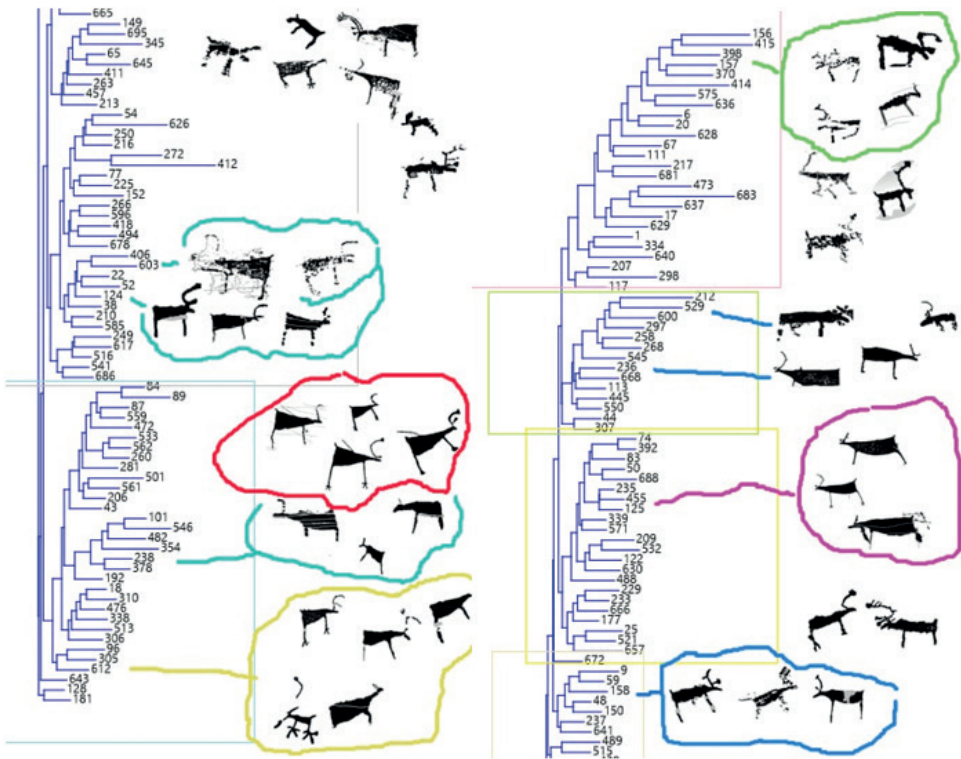
Илл. 31. Сетка деформаций эталонной формы по трем основным компонентам с использованием программы PAST.

³⁰ Официальная страница: URL: https://palaeo-electronica.org/2001_1/past/issue1_01.htm

- РС1(растяжение и сжатие формы вверх/вниз)– 38%.
 РС2 (закручивание формы по спирали)–18%.
 РС3(сжатие «спины» к центру, «брюшка» – от центра) –15%.

Последний вид анализа, который был применен к выборке,—это кластеризация методом NeighborJoining (в биологии это один из способов создания филогенетических деревьев происхождения видов), которая отразила примерно 14-17 групп сходных по форме выраженных кластеров, которые подлежат дальнейшей подробной интерпретации. Первичная интерпретация показала наглядные результаты разделения между собой отличных форм туловищ и конкатенацию схожих форм. Для удобства анализа данного кластерного дерева в будущем возможна разработка веб-интерфейса, включающего инвентарный номер, ID каждого из оленей, положение на местности (карта с географическими метками), уровень вложенности в кластерном дереве, а также интерактивное окно, в котором можно соотнести ID фигуры с изображением и увидеть другие изображения сходных с ним по форме фигур оленей из того же кластера.

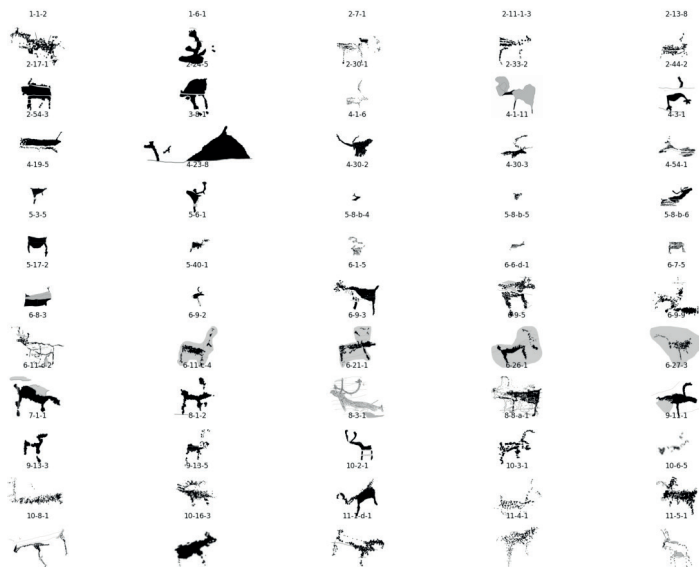
Предстоящий план работ над углублением анализа предполагает получение новых данных и интерпретацию результатов этого года работы во взаимосвязи с другими выделенными признаками фигур (бинарными, геометрическими), привлечение данных топографии (расположение фигур на местности—Чукотка, берег реки Пегтымель, горный массив Кэйныней—и корреляция расположения с кластерным анализом), дополнительные виды анализа (UMAP, геометрическая морфометрия на других структурных элементах фигуры оленя, применение модифицированного метода полуметок).



Илл. 32, 33. Частичная интерпретация кластерного дерева, полученного в программе PAST.

2.2. Латентное пространство признаков

Один из экспериментальных методов автоматизированного (и без обучения) получения признаков изображений—получение латентного пространства признаков обученной рисовать оленей GAN³¹. Интерес заключался в том, чтобы увидеть, какие различия будут между векторами и как они кластеризуются. Для пробы пера была выбрана базовая генеративно-состязательная нейросеть [10, с. 95–129] с двумя свёрточными слоями с 64 фильтрами размером 3×3 с шагом 2×2, реализованная с помощью библиотеки Keras, которая сначала успешно прошла апробацию на проверенном наборе данных (MNIST). Несмотря на очевидную разницу в объёмах обучаемой выборки (MNIST–70000, петроглифы–727), графичность и кажущаяся простота силуэтов оленей сопоставимы с чёрно-белыми рукописными цифрами датасета MNIST. Так, изображения оленей прошли первичную обработку, а также были преобразованы в одноканальные, развернуты все в одну сторону, переведены в белые силуэты на чёрном фоне и сохранены в формате 28×28 пикселей по аналогии с данными MNIST. Из общего датасета были удалены краевые случаи (Илл. 34): это изображения неполные и с трудноразличимым силуэтом. Таким образом, выборка для обучения состояла из 727 изображений и нейросеть³² быстро уходила в `modecollapse`³³. Предположительно это может быть связано с: а) маленьким размером обучаемой выборки, б) с «шумными» в связи с передачей выивки изображениями. Варианты лечения для последующей апробации: увеличение размера выборки с помощью дублированных, но слегка подвёрнутых изображений, увеличение разрешения изображений, обучение на наборе полученных контуров, а также использование нейросетей с альтернативной архитектурой (например, с автокодировщиком).



Илл. 34. Краевые случаи, удаленные из выборки для обучения нейросети.

31 Это метод машинного обучения, в ходе которого свёрточная нейросеть учится выявлять в выборке данных паттерны и генерировать на их основе изображения (и не только). В процессе обучения генеративная модель создаёт векторы из случайных чисел, которые в самом начале лишь шумные изображения, а дискриминативная пытается отличить реальные изображения от фейковых. Так, две модели параллельно тренируются, пока генеративная модель не начнёт генерировать правдоподобные изображения. По завершении с полученными признаками можно работать как с обычными векторами. [10, с. 7–12].

32 URL: <https://github.com/Petroglyphs/petroglyphs/blob/main/GAN.ipynb>

33 “It’s exciting because although the results achieved so far are significant...It’s frustrating because the models are fussy and prone to failure modes, even after all care is taken in the choice of model architecture, model configuration hyperparameters, and data preparation” [10].

Заключение

В результате работы с изображениями оленей с пегтымельских петроглифов были подготовлены наборы данных из обработанных первичных изображений и контуров и из таблиц для последующего анализа. Также были опробованы и оценены различные методы получения контуров и признаков оленей: собранные вручную, методом морфометрии. Работа с бинарными и геометрическими признаками методом «вручную» осложняется большим количеством пограничных случаев. По этой причине возникла необходимость включить в исследование альтернативные машинные методы получения признаков. Применение метода геометрической морфометрии показало наиболее репрезентативные результаты вычленения групп при проведении кластерного анализа методом neighborjoining, что позволяет углубить исследование в данном направлении. Потенциально перспективная линия получения признаков с помощью необучаемых нейросетей будет продолжена на следующих этапах работы и расширена. Полученные признаки в табличном виде позволяют обращаться и выделять различные группы изображений, что может быть использовано в построении базы данных, объединяющей в себе таблицы признаков и изображения.

Несмотря на высокое качество прорисовок методом протирки, участие художников в их получении накладывает отпечаток субъективной интерпретации на полученные изображения. Более того, среди прорисовок встречаются ошибки: например, некоторые трещины переданы в чёрном цвете и наоборот, какие-то детали выбитого изображения – серым. Это усложняет как машинную обработку, так и затрудняет интерпретацию исследователей. Избавиться от таких ограничений, а также увеличить выборку различных оленей возможно при получении прорисовок с созданных фотограмметрическим методом 3D-моделей поверхностей³⁴ с последующим автоматизированным выделением контуров хорошо обученным оператором. Решение этой задачи запланировано на следующем этапе работ.

Наши наработки впоследствии могут быть использованы как исследователями-археологами для подготовки данных, стилистического анализа других петроглифов, так и специалистами более широкого круга ввиду высокой адаптивности предлагаемой методологии к исследованиям. Так, получение контуров некачественных изображений может быть актуально в среде биологов.

Литература

1. Петроглифический отряд ИА РАН исследовал памятник наскального искусства на Чукотке // Российская Академия наук: сайт. URL: <http://www.ras.ru/news/shownews.aspx?id=89988e37-3ead-4ccb-a9b4-cdcfd8cfod&print=1> (дата обращения: 06.10.2022).
2. Дэвлет Е.Г., Миклашевич Е.А., Мухарева А.Н. Материалы к своду петроглифов Чукотки (изображения в скоплениях I–III на Кайкульском обрыве) // Изобразительные и технологические традиции в искусстве Северной и Центральной Азии. М.; Кемерово, 2012. Труды САИПИ. Вып. IX
3. Диков Н.Н. Наскальные загадки древней Чукотки. Петроглифы Пегтымеля. М., 1971.
4. Шер Я.А. Петроглифы средней и центральной Азии. М.: Наука, 1980.
5. Ватэ В. Возвращение к чукотским духам // Сибирские исторические исследования. – 2021. – №. 4. – С. 55-75.
6. Нормальное распределение // Википедия. URL: https://ru.wikipedia.org/wiki/Нормальное_распределение (дата обращения: 06.10.2022).
7. Васильев А. Г., Васильева И. А., Шкурихин А. О. Геометрическая морфометрия: от теории к практике. – Общество с ограниченной ответственностью Товарищество научных изданий КМК, 2018.

³⁴ Пример результата такого метода на петроглифах см.: [11].

8. Павлинов И.Я. Геометрическая морфометрия – новый аналитический подход к сравнению компьютерных образов // Информационные и телекоммуникационные ресурсы в зоологии и ботанике. СПб, 2001. С. 65-90. URL: <https://textarchive.ru/c-2940627.html> (дата обращения 05.10.2022).
9. Bookstein, F.L. Morphometric Tools for Landmark Data. Geometric and Biology. Cambridge University, Cambridge, 1991.
10. Brownlee J. Generative adversarial networks with python: deep learning generative models for image synthesis and image translation. – Machine Learning Mastery, 2019.
11. Свойский Ю. М., Леванова Е.С., Романенко Е.В., Конакова Е.С. Исследование моделей эстампажей амурских петроглифов из коллекции МАЭ РАН // Проблемы истории, филологии, культуры. М., Магнитогорск, Новосибирск: 2019. – №. 2 (64).

Приложение А (Цифровые решения)

Репозитории с кодами на языке Python:

1. [GAN](#)
2. [Markup](#)
3. [Первичная обработка](#)
4. [Получение контуров](#)

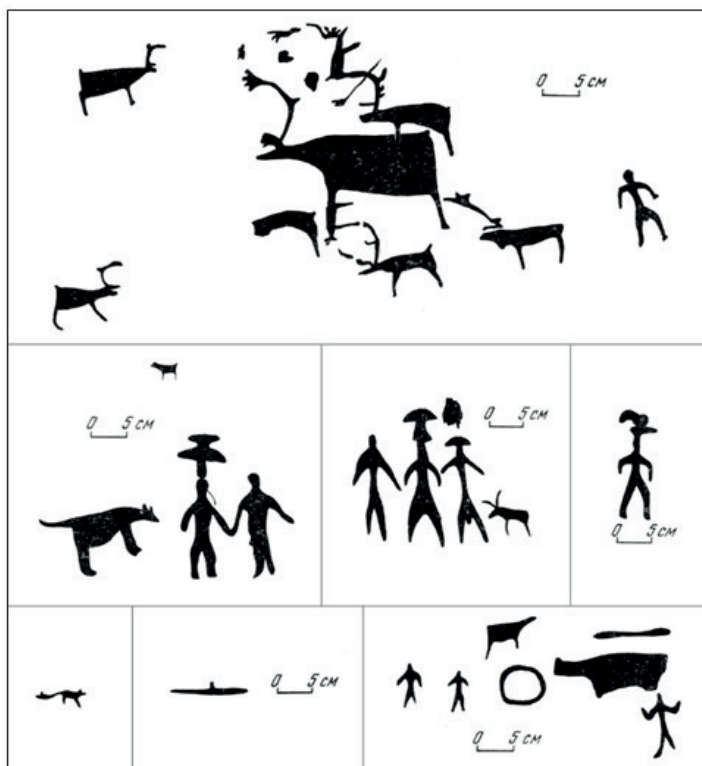
Инструменты Python:

1. Open CV
2. NumPy
3. Keras
4. Scikitimage

Программное обеспечение:

1. [tpsUtil](#), [tpsDig](#), [tpsRelw](#)
2. [PAST](#)
3. Paint
4. AdobePhotoshop

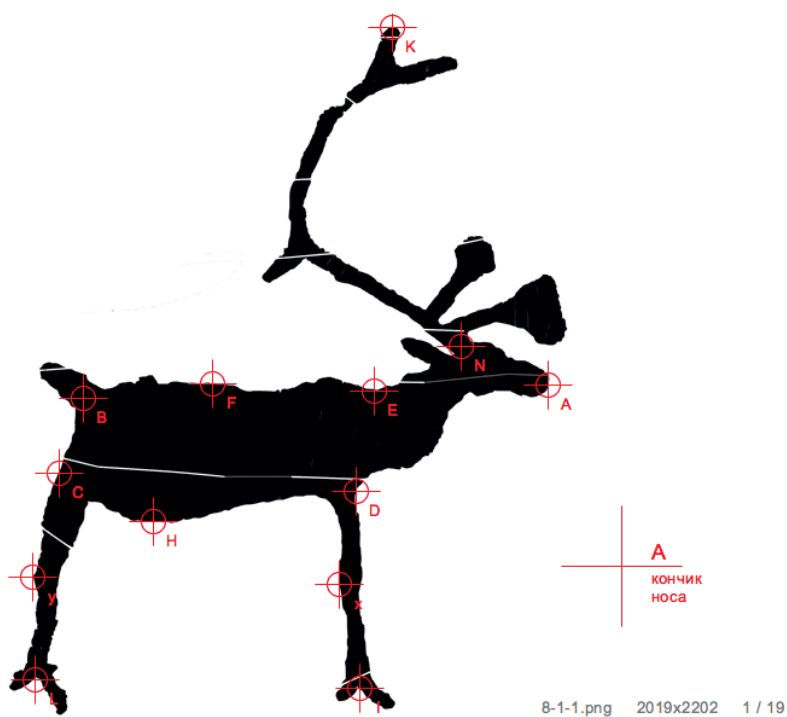
Приложение Б



Илл. Б-1. Прорисовки петроглифов скопления I по Н.Н. Дикову [3].



Илл. Б-2. Типы копыт



Илл. Б-3. Процесс нанесения меток координат в среде Python.

1	filename	Ax	Ay	Bx	By	Cx	Cy	Dx	Dy	Ex	Ey	Fx	Fy	Hx	Hy	Ix	Iy	Lx	Ly	Kx	Ky	Nx	Ny	xx	xу	yx	yy	
2	8-1-1.png	1710	1138	297	1178	224	1406	1127	1461	1182	1156	690	1134	510	1552	1138	2059	150	2033	1237	51	1446	1020	1075	1743	143	1721	
3	8-1-2.png	1340	762	382	810	358	1046	905	1058	921	930	645	850	572	1122	838	1456	324	1526	1385	193	1101	841					
4	8-2-1.png	407	2736	6014	3685	5866	4831	2440	4079	2477	3093	4313	3488	4264	4794	1762	5755	5780	7037	2428	518	1208	2613	2440	4720	5841	5718	
5	8-2-2.png	448	2021	4347	1722	4365	2536	1979	2506	1944	1764	3289	1776	3283	2739	2177	3295	4407	3175	1011	173	1118	1782	2129	2894	4443	2745	
6	8-2-3.png	1226	2202	2376	1980	2477	2274	1714	2327	1714	2100	2076	1989	2076	2293	1685	2723	2607	2689	304	183	1434	2028	1801	2549	2574	2458	
7	8-2-4.png	668	2044	2795	2158	2865	2610	1585	2572	1471	2088	2203	2114	2209	2814	1483	3444	2960	3489	598	350	1031	1961					
8	8-3-1.png	785	744	1773	885	1779	1032	1341	1047	1304	716	1560	900	1566	1054	1344	1516	1792	1520	678	166	875	675					
9	8-4-1.png	1574	386	210	532	208	785	999	678	1021	376	573	469	600	780	1089	910	247	1101	1153	16	1249	382					
10	8-5-a-1.png	156	727	1617	890	1630	1155	724	954	731	753	1177	826	1171	1168	683	1605	1512	1764	204	83	405	651					
11	8-5-b-c-1.png	2258	785	177	372	189	915	1612	861	1612	650	915	530	908	1006	1690	1501	218	1653	1624	50	1873	694	1542	1205			
12	8-6-1.png	855	339	105	148	80	305	630	378	679	274	419	202	352	388	601	485	55	569	874	43	721	279					
13	8-7-1.png	852	160	154	132	204	303	697	261	718	165	442	150	293	381	789	497	198	617			782	148	687	384		265	1292
14	8-7-2.png	1484	527	223	678	265	966	1207	731	1187	549	719	617	605	983	1319	1051	308	1399	1192	39	1238	515			265	1292	
15	8-8-a-1.png	448	267	74	222	89	305	288	261	300	207	201	191	200	290	295	380	88	454	499	74	339	169	257	311	66	397	
16	8-8b-1-1.png	1061	524	157	492	201	648	772	557	793	420	462	422	470	659	759	880	256	1013	1011	93	922	417					
17	8-8b-1-2.png	478	663	1240	101	1302	283	868	438	714	392	978	229	1036	407			1396	547	82	334	624	392					
18	8-8b-1-3.png	236	508	1055	358	1074	499	583	483	594	357	848	341	859	547	523	762	1001	817	40	107	414	376					
19	8-8b-1-4.png	787	407	159	327	193	479	596	458	636	357	400	293	355	572	606	729	254	778	486	28	685	358					
20	8-8b-1-5.png	1304	784	172	421	206	759	934	775	979	629	558	477	399	853	894	1187	175	1178	1458	67	1147	650					

Илл. Б-4. CSV-файл с координатами меток.