NATIONAL
ENDOWMENT
FOR THE
HUMANITIES

OFFICE OF **DIGITAL HUMANITIES**

Narrative Section of a Successful Application

The attached document contains the grant narrative and selected portions of a previously funded grant application. It is not intended to serve as a model, but to give you a sense of how a successful application may be crafted. Every successful application is different, and each applicant is urged to prepare a proposal that reflects its unique project and aspirations. Program guidelines also change and the samples may not match exactly what is now required. Please use the current set of application instructions to prepare your application.

Prospective applicants should consult the current Office of Digital Humanities program application guidelines at https://www.neh.gov/grants/odh/digital-humanities-advancement-grants for instructions.

Applicants are also strongly encouraged to consult with the NEH Office of Digital Humanities staff well before a grant deadline.

Note: The attachment only contains the grant narrative and selected portions, not the entire funded application. In addition, certain portions may have been redacted to protect the privacy interests of an individual and/or to protect confidential commercial and financial information and/or to protect copyrighted materials.

Project Title: *SnowVision: A Machine Learning-Based Image Processing Tool for the Study of Archaeological Collections*

Institution: South Carolina Department of Natural Resources

Project Directors: Karen Yvonne Smith, Song Wang, Jun Zhou, and Colin Wilder

Grant Program: Digital Humanities Advancement Grants, Level III

LIST OF PARTICIPANTS

Project Staff

Smith, Karen Y.	South Carolina Department of Natural Resources
Lu, Yuhang	University of South Carolina
McDorman, Samantha	South Carolina Department of Natural Resources
Miyasaki, Sydney J.	University of South Carolina
Wang, Song	University of South Carolina
Wilder, Colin	University of South Carolina
Zhou, Jun	University of South Carolina

Advisory Committee

Arnold, Taylor	University of Richmond
Bergel, Giles	University of Oxford
Carroll, Norine	University of West Florida
Carter, Andrew	University of West Georgia
Compton, Matthew	Georgia Southern University
Elliott, Daniel T.	Lamar Institute, Inc.
Gougeon, Ramie	University of West Florida
Keith, Scot	New South Associates, Inc.
Saunders, Rebecca	Louisiana State University
Semon, Anna M.	American Museum of Natural History
Shanks, Jeffery	Southeastern Archaeological Center
Snow, Frankie	South Georgia State College
Stephenson, Keith	University of South Carolina
Thompson, Amanda	University of Georgia
White, Nancy	University of South Florida

Additional Letters of Commitment

Sagano, Paul	University of South Carolina
Simmons, Matthew J.	University of South Carolina

Letters of Support

Galle, Jillian E.	Digital Archaeological Archive of Comparative Slavery
Kiernan, Kevin	University of Kentucky

NARRATIVE

1. Enhancing the humanities

This proposal outlines an initiative to implement *SnowVision*, a unique, machine learning-based image processing tool, and to disseminate this tool to scholars working in archaeological laboratories across the southeastern US. The *SnowVision* project, now in its third year, represents a successful collaboration among archaeologists, computer vision scientists, high-performance computer scientists, and digital humanities scholars. Over the last three years, our multi-disciplinary team has developed a one-of-a-kind image processing tool that works with a corpus of 1,500-year-old designs and design fragments from carved wooden paddles, including those shown in Figure 1. Paddle designs such as these have been preserved only as fragmentary and overlapping impressions on stamped pottery sherds. Painstaking human labor over the decades has resulted in the partial or complete reconstruction of approximately 900 paddle designs from the pottery sherds on which they were impressed (Broyles 1968; Snow 2003). The uniqueness of each design presents the rare opportunity to measure human social processes—such as migration, residential mobility, and interregional interaction—when sherd-to-design matches are discovered to occur on archaeological sites, as they often are, separated by tens or even hundreds of miles.

The aim of *SnowVision* is to automate the matching process between sherd and design, expediently and accurately matching fragmentary stamped pottery sherds from the archaeological record with their appropriate complete paddle design from among the hundreds of registered design reconstructions in the corpus. As demonstrated by our publications to date, *SnowVision* software is a success (Lu et al. 2018; Zhou et al. 2017), but challenges remain. First, our unique software is not currently available outside of the University of South Carolina (USC) campus where it is housed. Second, only sherd data necessary for initial software development has been digitized, meaning the extant sherd database is of limited scope. NEH implementation-level funding ($323,668) is sought to increase accessibility and expand digitization efforts.

To address accessibility, NEH funding will support the integration of *SnowVision* with the sherd and design database that contains geographically and contextually situated data and deliver both through our project website www.worldengraved.org. To scale-up the project, NEH funding will enable the USC team to assemble a 15-person advisory committee and will provide five key laboratories with start-up money to begin to digitize un- and under-studied collections stored at their home institutions. Funding five *SnowVision* satellite labs will allow the project team to begin creating a critical mass of digital data on sherd-to-design matches that will be made available through World Engraved and will assist the USC lead developers with needed offsite testing of the system. NEH funding also will enable ongoing refinement of the *SnowVision* application resulting from user testing, evaluation, and feedback.

In sum, we seek Level III support to 1) integrate *SnowVision* with an interactive, online user interface, 2) acquire and integrate feedback from scholars working in laboratories and curation facilities across the Southeast, 3) enhance the technological infrastructure of *SnowVision* so that the newly integrated system meets the needs of the user community and has a framework built for long-term success, and 4) assist select users with start-up funds to begin digitizing collections, providing the USC team with rigorous, off-site testing of the system.

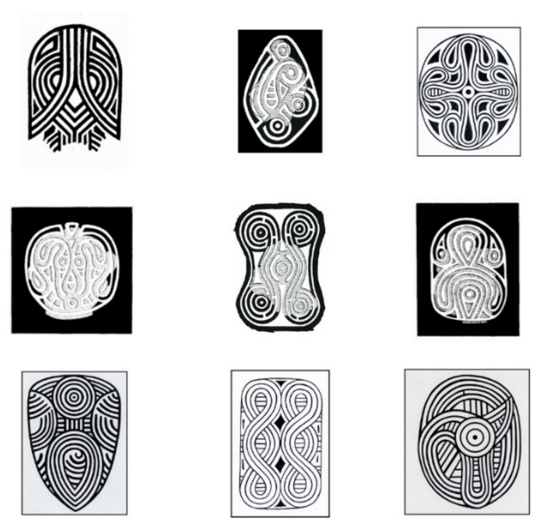


Figure 1. Paddle design reconstructions by Broyles (Row 1-#2, Row 2-all), Snow (Row 1-#3, Row 3-all), and an unknown illustrator (Row 1-#1).

1.1. Significance of subject matter

The archaeological record is filled with fragmentary objects of bone, pottery, shell, stone, wood, and cloth variously embellished with realistic and abstract designs. These designs may include figural imagery such as that seen on ancient Maya and Greek pottery vessels or the carved marine shell gorgets of late prehistory in North America. They may also include geometric designs such as those found on Ancestral Pueblo wares, carved bone pins from the American Southeast and ancient India, and carved bone pendants and gaming pieces from Newfoundland. Such imagery also includes maker's marks and seals placed on manufactured objects. The smaller the design fragment preserved on an archaeological object or the more diverse the design corpus, the more difficult it can be to match a fragment to a complete composition. As a result, large numbers of decorated objects found in the archaeological record have not contributed to broader studies relating to style, contexts of production and use, and meaning. *SnowVision* addresses this gap for an important class of artifacts.

Elaborately carved wooden paddles of the Southeastern Woodlands represent an ancient Native American art form, one with rules of stylistic design and technical execution that were taught in communities of learning and passed on from one generation to next for nearly twenty centuries. Despite the artistic tradition's longevity, these carved paddles rarely receive mention in either scholarly or popular tomes on primitive art (e.g., Feest 1980; Roosevelt and Smith 1979; Wilson 1984), in part, because the paddle designs are imperfectly preserved as impressions on pottery vessels, which are themselves recovered from the archaeological record in fragmentary rather than whole form. The wooden paddles have not survived, and so our knowledge of them comes only from modern artistic reconstructions made piece by piece from sherd fragments.

The initial curvilinear expression of the craft, called Swift Creek after the archaeological site on which it was first identified, dates between AD 1 and 800. During this period, highly creative and complex curvilinear paddle designs were not the exception but the rule. As Robert Wauchope (1966:55) aptly noted "this degree of experimentation…was never equaled in southern prehistoric art before or since." We find the most ornate paddle impressions on hundreds of thousands of pottery sherds of the Swift Creek style tradition made during the 300-year Middle Swift Creek period (ca. AD 350 - 650) when the paddle craft was at its most diverse and creative. Designs dating to Middle Swift Creek have provided material for the development of *SnowVision* but our technical methodology can be applied to carved paddle designs from any subset of the paddle craft tradition, including those Mississippian Period paddle designs that post-date AD 1000. In short, *SnowVision* has the potential to impact Native American research *across a two-thousand-year period.*

Some 1,820 archaeological sites with Swift Creek (AD 1-800) components have been formally recorded across the present-day states of Alabama, Florida, Georgia, and South Carolina (Smith and Stephenson 2017). Hundreds of archaeological site collections and hundreds of thousands of Swift Creek pottery sherds from these sites reside in curation facilities and laboratories across the south (Table 1). For the past three years, our multi-disciplinary project team has focused on developing the matching and image processing algorithms that reside at the heart of *SnowVision*. With NEH support, we seek to bring scholars and curators from some of the key institutions listed below in Table 1 together to engage with *SnowVision* and with each other and the project team. User input is critical for the integration of the software with an online user interface and the database that houses sherd and design metadata.

We also request funding to provide five curation facilities with $5,500 each to function as satellite labs for the duration of the project. As many facilities listed in Table 1 are located on or near university campuses, *SnowVision* satellite labs will be able to tap into a pool of locally available student labor to perform day-to-day scanning and metadata entry. This is a cost-effective way to expand the database, while also providing students with work experience and labs with funding not otherwise available. Critical to this grant request is the direct feedback the USC team will receive from these satellite labs in conjunction with the project. These are not simply digitization efforts. Rather, they represent efforts directed at testing the usability of the system outside of the USC campus over a prolonged period. Funding of $5,500 is sufficient to employee one student on a part-time basis for two semesters. Five facilities were chosen for the significance of their holdings to the research themes outlined below. For example, the University of West Georgia houses the Leake Site collection. Designs in this collection have never been studied at the sherd level, and yet the Leake site was a major gateway for the exchange of goods and information between the Deep South and the Midwest from AD 300-500.

Table 1. Select curation facilities with important Swift Creek collections. Highlighted rows indicate proposed *SnowVision* satellite labs.

Curation Facility	Location	Select Swift Creek Site Collections
Fort Benning Military Installation	Columbus, GA	Quartermaster
Georgia Southern University	Statesboro, GA	Shelly, Hartford, Solomon, Milamo
Museum of Science and History	Jacksonville, FL	Dent, Mayport
Division of Historical Resources	Tallahassee, FL	Letchworth, Block-Sterns
Office of Archaeological Research	Moundville, AL	Simpson's Field
University of Florida	Gainesville, FL	McKeithen, Carrabelle
University of Georgia	Athens, GA	Mandeville, Swift Creek, Fairchild's Landing, Kolomoki
University of South Florida	Tampa, FL	Otis Hare, Overgrown Road
University of West Georgia	Carrollton, GA	Leake, Kolomoki
University of West Florida	Pensacola, FL	Bernath

1.2. Research themes

By funding this proposal, NEH will be supporting the integration of software with an online interface that will enable scholars to explore a diverse set of research themes, without the lag that accompanies the painstaking search for matches that has traditionally plagued Swift Creek researchers. The subject matter warrants such attention: Swift Creek designs and the vessels on which they were impressed have exceptional research potential.

Figure 2. Swift Creek design (right) reconstructed from fragmentary pottery sherds (left). Photo and design reconstruction by Frankie Snow.

As early as the 1890s, archaeologists began to realize that 1) complete paddle designs in the Swift Creek style could be reconstructed from fragmentary impressions on sherds, as in the example on the right (Figure 2), and 2) certain paddle designs were widely distributed across the lower Southeast, sometimes separated by hundreds of miles (Broyles 1968; Holmes 1894; Snow 1975, 1982, 1998; Snow and Stephenson 1998; Wallis and O'Dell 2011). From these two realizations were borne interesting, novel research possibilities that Swift Creek scholars have sought to address ever since.

1.2.1. Residential mobility, migration, and interaction

Traditional archaeology has met with variable success monitoring mobility and related scales of human movement over space. One difficulty, among others (c.f., Berelov 2006), derives from our inability to make inferences about individuals in the remains of stone, bone, and pottery—common materials on archaeological sites in the Southeast. Such remains more often represent, effectively, a palimpsest of individuals and groups. How often do people move about on the landscape, how far do they move, and for what purposes? These are questions that all-too-often evade archaeologists because individual movements remain elusive.

Swift Creek paddle designs provide evidence for the individual that can overcome this deficiency in the archaeology record, opening a unique window on the past. For example, Snow (1977) noted a high number of Swift Creek design matches among sites within the big bend area of the Ocmulgee River (called the Ocmulgee-Big Bend) and between the Ocmulgee-Big Bend and the headwaters of the Satilla River to the south. These matches occurred within an approximately 30-mile (48-kilometer) radius, providing, as Snow said, clues to settlement patterns (i.e., mobility) within the region. He also noted a more limited number of design matches between those sites and sites much further to the northwest, southwest, and southeast, well outside of the Ocmulgee-Big Bend region. These longer-distance, inter-regional, matches suggest another kind of social process at play (e.g., long-distance foraging or interregional aggregation). An example of the design connections between a site (Milamo/9WL1) and its inter- and intra-regional neighbors is shown in Figure 3 (right).

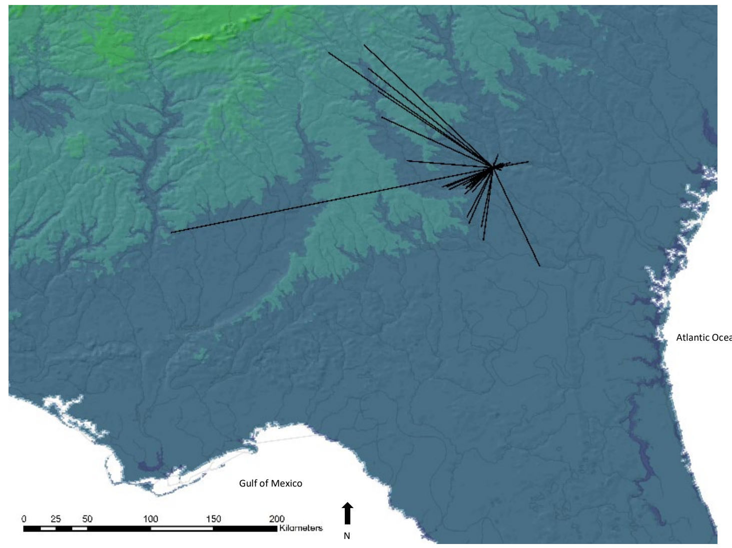


Figure 3. Design matches between Milamo (9WL1) and other sites in the region. Black lines represent one or more paddle design matches. Most matches with Milamo occur within the Ocmulgee-Big Bend (Stephenson et al.2012).

Studying design connections across contemporary household middens within a village site also holds research promise. Saunders (1986) examined Swift Creek designs in individual household middens at the Kings Bay site, Georgia. Snow and Stephenson (personal communication 2015) studied designs among individual middens at the Hartford site, Georgia. Both Saunders (1986) and Snow and Stephenson identified paddle design connections that linked each household trash midden to one or more middens across their respective study sites. Smith and Knight (2017:127) have taken these patterns to indicate that paddle production and use operated independently of pottery discard, but further data is needed to test this idea. *SnowVision* will facilitate such data collection and exploration. Furthermore, *SnowVision* can effectively match sherd designs to other sherds even in the absence of known design reconstructions. Future work will pursue algorithms to automatically create new designs from the sherd data we have collected, expanding the design corpus by several fold.

Using instrumental neutron activation analysis (INAA) to source ceramic vessels and local clays, Wallis (2011) documented instances in which Swift Creek vessels themselves apparently moved great distances. Among the vessels he analyzed, Wallis (2011:114) found that Swift Creek vessels from mortuary mounds on the lower St Johns River, on the Atlantic Coast of Florida, were made with non-local clays; whereas, Swift Creek vessels from residential sites in the same area were made with local clays. Many vessels of nonlocal origin may have come from sites much further to the north, up the Atlantic Coast. His argument about the long-distance movement of pots found in mortuary contexts was bolstered by independent evidence related to paddle design matches among sites. Expediting our ability to discern paddle design matches among sites, as *SnowVision* does, will enhance research that addresses how foragers organized and moved around on the landscape 1,500 years ago.

1.2.2. Indexes and iconography

Broyles (1968:52) was one of the first archaeologists to suggest that some Swift Creek designs may represent animals or plants, though most designs are decidedly abstract. The idea that a subset of Swift Creek designs within the total corpus may carry symbolic meaning was further explored and expanded on by Snow (1998), inspiring some Swift Creek researchers to go in a direction distinct from that related strictly to the movement of people. In a more nuanced stance on meaning, some have argued that Swift Creek paddles were indexed to people or things, rather than being direct symbols of plants, animals, and people.

For example, Wallis (2013:211; see also Wallis 2011) argued that Swift Creek had indexical properties which served as a form of maker's mark linking the wooden paddle to the vessels on which they were stamped and to the individuals who performed the stamping. Taking the argument even further, Wallis (2013:213) proposed that Swift Creek vessels served as distributed objects connecting multiple places and persons simultaneously. Regardless of whether Swift Creek paddles and the vessels on which they were impressed were representational or indexical or something else entirely, pursuit of such research themes will benefit from the integration and scaling-up of *SnowVision*.

1.2.3. Artistic style and creativity

Recent research into Swift Creek designs suggests that Swift Creek artisans were deeply concerned with creating unique paddles during the Middle Woodland period (Smith and Knight 2017). Critical to this argument is our understanding of the steps and stages, or tasks, involved in creating a design on the face of a paddle. Smith and Knight (2014:128) have suggested that it is only through a study of the production sequences of whole designs that scholars can begin to define the style formally, delineate its structural elements and their possible combinations, delineate specific subtraditions or stylistic "lineages," and distinguish between broad geographic style areas using a defined set of core elements, such as those identified in Smith and Knight (2014).

These facets of style research are not ends to themselves. Rather, they inform broader theoretical interests, such as how long-lived cultural traditions, which archaeologists are uniquely positioned to study, developed and were maintained (e.g., Tehrani and Riede 2008). Investing time and effort in the development and delivery software that advances the goal of design recognition on individual sherds is important to the above and to our future project goals. Such efforts will provide a software baseline for future development related to whole design reconstruction made possible only through sherd-to-sherd matching.

1.3. Transforming research

As we have argued above, each of these diverse research programs will benefit when researchers are given the tools to automatically search and find matches within the Swift Creek design corpus and to download related sherd and design metadata. No longer will scholars be confined to research small areas of the Swift Creek world that have been painstakingly studied by a few individuals. *SnowVision* will open up research opportunities across the entire region and throughout the entire two-thousand-year period during which carved pottery paddles were in vogue. Our plan to simultaneously deliver detailed spatial and contextual data on those designs and the sherds that bear their impressions extends the research value of *SnowVision* even further.

SnowVision, however, has the potential not only to enhance current research programs, such as those mentioned above, but also to transform them. Untapped research that *SnowVision* will make possible includes studying the position or rotation of the paddle relative to the vessel. From this, archaeologists can begin to infer the handedness of potters (Sassaman and Rudolphi 2001; Vidal 2017). Our computer vision software also can assist in the classification of paddle designs. Classification helps the archaeologist sort designs by time period. Classification also has the potential to identify individual paddle makers or communities of paddle makers that share the same stylistic conventions and, by inference, are members of the same learning or practice groups. Although classification and handedness studies could be done manually, the use of *SnowVision* software will not only expedite such inquiries across a large dataset but will also add an unbiased element to the classification process. Identifying design communities through classification and handedness studies is a novel area of Swift Creek research that *SnowVision* will make both possible and, for the first time, feasible.

SnowVision has the potential to contribute to a number of additional research domains. The most-closely related domain concerns Mississippian period (AD 1100-1500) paddle designs that, though quite different from Swift Creek designs, are nevertheless part of the same broad technological tradition. Extending *SnowVision* to this material would only require expanding the database to include those later designs and their metadata. Ms. Semon brings her knowledge of this later material to the advisory committee. Applications outside the field of archaeology include images and other content in the fields of numismatics and woodblock printing, among others. Our curve-matching method would likely need to be adapted to be applicable to these new materials. However, our approach to the classification and matching of curved lines could offer researchers in parallel fields novel solutions for matching different instances of the same or similar visual content (see Appendix E). Feature matching, which is a common method used in the fields mentioned above, results in information loss. Our method is more accurate than feature matching because it provides an exact match at the image level, making it a more attractive approach depending on the research questions.

2. Environmental scan

Computer vision technologies increasingly have been used to foster new insights into humanities and social sciences subjects. In this section, we illustrate several examples of similar computer vision projects and then discuss how our work differs from them.

Implemented within the PENED (Reinforcement Programme of Human Research Manpower) project, Democritus University of Thrace has developed a web-based 3D pottery search engine using a pottery sherd's surface curves and its plane orthogonal to the axis of symmetry (Koutsoudis et al. 2010). The Pottery Goes Digital project has developed an identification pipeline from 3D scanning of pottery sherds to sherd edge analysis and the pottery making processes (Lami et al. 2015). The SHAPE Lab - Stitch project has provided semi-automatic methods for reconstructing pottery, murals, and sculptures from fragments (Cooper et al. 2001). The Stanford Forma Urbis Romae project provided methods and tools to search for matches in an ancient marble map of Rome (Koller and Levoy 2006). This project assembled over 1,186 fragments of Rome maps that dated back to the early 3rd c. CE.

The Theron Wall Paintings Reassembling project has implemented a system for high-volume acquisition and matching of fresco fragments (Shin et al. 2012). The 3D Colonial Philadelphia Project has studied virtual reconstruction of thin-shell archaeological vessels using expert priors and surface markings (Cohen et al. 2010). It drew upon the ceramic assemblage recovered during recent historical archaeological research on the Mall in Independence National Historical Park, Philadelphia.

Our project differs from these projects in that our work concerns pattern patching from the partial sherd pattern to the full design pattern. Our work also differs from the ones discussed above in that there may be overlapping copies of a paddle pattern present on a sherd, what archaeologists commonly refer to as overstamping. *SnowVision* identifies these copies, decomposes them, and then matches them individually to the whole design pattern.

3. History of the project

Since our initial meeting in August 2015, the USC project team has worked to develop and refine the algorithms that drive *SnowVision*. From July 2016 through May 2017, *SnowVision* was funded by start-up grants from USC ($20,000) and the National Parks Service ($39,848). This funding supported work in two project domains: data acquisition and prototyping computational methods. With this funding, Dr. Wang, Ms.

Zhou, and students tested and selected the best matching metric and then developed and refined the pattern decomposition algorithm. Project code generated during the above grant period has been published on GitHub (https://github.com/*SnowVision*/SnowVision1.0) and documented in a peer-reviewed publication in the *Journal of Electronic Imaging* (Zhou et al. 2017).

With a working prototype in-hand, the team secured a two-year grant from the National Science Foundation ($235,487, NSF Grant #1658987) for the period from August 2017 to April 2019. NSF funding pays a portion of Dr. Smith's salary as well as two years of tuition and stipend for an archaeology graduate student. Smith and the student are continuing to collect sherd image data and metadata, including additional sherd data acquisition at Georgia Southern University (GSU), which houses the collections related to Frankie Snow's design work, and the University of Georgia (UGA), which houses the collections related to Bettye Broyles's design work. Both Snow's and Broyles's reference collections are crucial to the ongoing refinement of *SnowVision*. The NSF grant pays tuition and stipends for Dr. Wang's students and one month of Dr. Wang's summer salary in 2018 to further research algorithm development and image processing. NSF funding has also supported the development of a draft website (www.worldengraved.org) and a working database for storing sherd, design, and image metadata. Website wireframes and *SnowVision* technical details are provided in Appendix C and D, respectively.

To date, the team has completed the digitization (3D scanning) of 3,000 pottery sherds curated at Georgia Southern University. These new data have allowed the successful completion of an automatic method to extract curves on sherds and the development of an automatic matching method to identify underlying designs on sherds with single and composite or "overstamped" patterns. The USC team has also created a framework for storing design, sherd, and image metadata and constructed a draft website for delivering *SnowVision* to the user community. Our most-recent computer vision work has been accepted into the Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence (Lu et al. 2018).

As mentioned above, current NSF grant work supports further algorithm development, such as optimizing the deep learning neural network for better recognition accuracy, fewer network parameters (i.e., more compact network architecture) and lower computational complexity (i.e., faster algorithm). In addition, work to be completed in the coming months includes 1) developing a protocol (based on NextEngine 3D scanner) to collect the depth image of sherds, 2) collecting additional depth images from sherds housed at GSU and UGA, and 3) developing an effective algorithm in segmenting the curve structures from the depth image of sherds.

If funded, NEH grant work will focus on two new project domains: implementation and delivery of *SnowVision*. The outcome of our project will be an integrated digital resource that serves to advance our understanding of southeastern Native American paddle designs and the people who made and used them. As shown in Figure 4 (below) and described below, four complementary components make up this digital resource: (a) 3D pottery sherd digitization, (b) the World Engraved website developed in a Django framework and hosted on Google Cloud Platform, (c) the collection database in PostgreSQL, and (d) the backend *SnowVision* server housed on an XSEDE resource. The database and website are integrated at this time. NEH funding will support the integration and delivery of all four components.

With NEH support, we will host and manage the WorldEngraved website on a Google Cloud Platform (GCP) infrastructure. The flexibility of cloud-based hosting will be a boon for the sustainability of the project. Downtime/the site being offline will be minimized to industry-leading levels, thus ensuring users will, for all intents and purposes, always be able to access the site. Cloud computing also allows for flexibility in compute resources; that is, processor cores and memory can be added on-the-fly according to user demand and project needs. As such, the site will be able to respond to spikes in demand for access—something we hope for as the project grows in influence and popularity—without taxing the server. This will help to ensure uptime and responsiveness, improving the overall user experience.

(a) (b) (c)

Figure 4. World Engraved framework: (a) 3D pottery sherds digitization, (b) World Engraved website, (c) World Engraved collection database and (d) the backend SnowVision server.

In addition to processing tens of thousands of sherd images collected or contributed from different resources, we need to perform matching queries between all possible pairs of sherd and designs and match each pair. That is, we have to exhaustively search each possible pose for candidate pattern patches and further apply Convolutional Neural Network (CNN) to identify these patches. This results in a large amount of computation time, which cannot be effectively satisfied by a single or few CPUs. For example, based on our study, performing one single sherd-to-design matching over a design database of 98 takes roughly half an hour. We will leverage high-performance computing (HPC) facilities to speed up the large-scale matching task. In particular, we will use both the parallel job schedulers (e.g. SLURM) and machine learning packages (e.g. Caffe, Tensorflow) on GPU to improve the performance. To serve this purpose, we will request HPC resources from the Extreme Science and Engineering Discovery Environment (XSEDE), the national NSF-funded virtual system that share resources like supercomputing, data and tools. XSEDE provides a science gateway, or a set of tools that help assemble resources like HPC computing, scientific applications and data, and build automatic workflows via a portal, usually in a graphical user interface. We will request and customize a XSEDE science gateway to integrate our WorldEngraved website with the backend *SnowVision* matching algorithms running on XSEDE HPC clusters. A proposal requesting 1-year-startup resources to XSEDE has been approved, and now is utilized by our research team. Should our NEH proposal be approved, we will further request advanced resources at the research level from XSEDE to continue supporting our *SnowVision* project. Implementing the *SnowVision* project on a XSEDE platform can foster collaborations among researchers across domains, and most of all, be cost-effective.

Funds from the grant will be used to help pay for the project's share of website hosting at an estimated cost of $3,500 plus the cost of a student system administrator to run the website. The estimation is based on 1 1TB GCP virtual machine with 4GB memory for the course of 3 years. The Center for Digital Humanities (CDH) at University of South Carolina, led by Dr. Matthew J. Simmons, will commit to supporting the project—including security updates and maintaining the site on our GCP infrastructure—for 5 years, or the length of the grant, whichever is longer, up to 7 years. CDH is willing to continue to support the site after that time period. However, as internet and web technologies change so rapidly and regularly, predicting how long-term our support can be is impossible; a website built in 2019 may not function on the standard technologies of 2030. An additional $7,791 per year will be used to pay for one student developer to integrate the WorldEngraved website into XSEDE science gateway. The Research Computing (RC) group at University of South Carolina, led by Mr. Paul Sagano, will commit to supporting the project—including collaborating with XSEDE, administrating the RC student developer and supporting additional HPC resources —for the length of the grant.

4. Work plan

September – December 2019

Dr. Smith will meet with project staff in September to review project goals, milestones, and the work plan presented below. Staff will also discuss plans for bringing the advisory committee to Columbia, SC, including proposed meeting dates. From this discussion, Dr. Smith will begin making preparations for the committee's visit. One risk that might delay the project is the difficulty of finding a single workable meeting date for fifteen advisory committee members. To keep on schedule, we will offer a robust virtual meeting option for anyone who cannot make the meeting in person. However, we will aim for a single, in-person date that works for the majority, utilizing a doodle poll. Dr. Wang, Ms. Zhou, and the computer vision team will initiate the XSEDE HPC cluster during this period. Ms. Zhou will test and tune the cluster to achieve the best performance status within two months. The XSEDE student web developer also will be hired during this period and will begin working with Mr. Miyasaki and the team to achieve integration.

January – March 2020

During this period, the project team will meet face-to-face with the advisory committee. We will review project achievements and discuss project plans, both long and short-term. At the meeting, we will provide demonstrations of all stages of *SnowVision* and discuss in detail how the committee views the usability of the system. Since the advisory committee is comprised of several potential contributors, the question of licensing content will be an important discussion point at the meeting. Based on our experience we believe that the Creative Commons Attribution-NonCommercial-ShareAlike (CC BY-NC-SA) will best suit the needs and wishes of our individual and institutional contributors. One risk during this period that might delay the project is if staff and participants do not agree on a one-size-fits-all solution to licensing of contributed content. In this case, staff will a customize solution so that individual and institutional contributors unique licensing wishes are met. Also, during this period, Mr. Lu, working with Dr. Wang and others on the team, will continue to refine the software to improve accuracy and speed.

April – July 2020

With feedback from the advisory committee, the project team will make refinements to the online platform, user workflows, and licensing protocols. The team also will continue to work on developing and testing the infrastructure. Ms. Zhou and Mr. Lu will work with CDH and RC web and XSEDE student developers to integrate the software with the website. Near the end of the period, an advisory committee meeting will be held remotely via a platform such as Google Hangouts, when the committee will be given an update on the integration of the website with the matching software and plan for the first phase of testing. The team will initiate a (limited) soft launch of the website in advance of this phase. Uploading of new database content will be restricted to advisory committee members and project staff. Ms. McDorman will review metadata entry and field data entry questions from users. One risk during this period is that one or more committee members will be unable to assist with testing. However, we think this risk is extremely low given the high level of interest among committee members, as their letters of commitment attest, in helping bring *SnowVision* to full implementation. Should any one member drop out for an unanticipated reason, Dr. Smith has several excellent alternates in mind.

August – December 2020

During this period, all committee members with access to Swift Creek collections will be asked to create metadata and upload sherd image data for routing to the image matching software. Many of our committee members have access to 3D scanners through their home institutions. For those who do not, we will offer to scan their test sherds for the project. It is also possible that scanning sessions can be organized to coincide with one of several national society meetings that many on the advisory committee attend. Also, during this period, *SnowVision* satellite labs will be in full operation, contributing new content through the World Engraved website. Any new sherd content submitted during this period will be processed through *SnowVision* for matching outcomes, and any true matches will be noted in the database with results delivered to the users. These sherds and their matching results will become a permanent part of the database.

January – May 2021

Satellite labs will continue to contribute content and test the system. Ms. McDorman will check the metadata for obvious errors or issues and work with labs to correct them. She will also work with Dr. Smith to make refinements to the data entry and scanning protocols, as needed, and work with Mr. Miyasaki to update those on the website. Ms. McDoman will send out an evaluation survey to stake holders (see Appendix E).

June – August 2021

During the final months of the grant, the team will finalize website and software content. Ms. Zhou and Mr. Lu will upload all current code to GitHub. Dr. Smith and Dr. Wilder will prepare an announcement about the product for distribution through appropriate archaeology and digital humanities channels. Dr. Smith, working with senior staff, will draft a white paper for submission to NEH, documenting project achievements and lessons learned.

5. Final product and dissemination

In the final months of the grant, we will initiate a hard launch of the website www.worldengraved.org (currently in draft form) that will deliver Swift Creek designs, sherd images, metadata, and search tools including the GUI for *SnowVision* to scholars and the public. We will work with KUDOS ([https://grow kudos.com](https://growkudos.com)), a free service for researchers, to disseminate information about our project to a broad audience. Through the project website, we will provide written guidance to users on 3D image acquisition and guidelines for metadata creation and use. We will upload all current project code to GitHub for adaptation to other subject matter. We will produce a white paper detailing project activities and scholarly response to the software. Any new computer vision advances made during the project will be shared through scholarly articles. Where appropriate and possible, all website products will conform to Web Content Accessibility Guidelines (WCAG) 2.0 and Section 508 Standards for Electronic and Information Technology of the United States Workforce Rehabilitation Act of 1973, as amended.

Through the formation of an advisory committee, our project will have secured the buy-in from major figures in Swift Creek pottery studies. Their participation will encourage downstream use of our program at other institutions with whom they have research relationships as well as at their home institutions. The addition of advisory council members Dr. Arnold and Dr. Bergel, who have research interests and extensive digital humanities experience in adjacent disciplines, will help ensure the cross-pollination of our work with theirs and the extensibility of *SnowVision*'s in the long term.

Further, we believe that the implementation of *SnowVision* will motivate scholars across the Southeast to digitize archaeological specimens that reside in boxes on curation shelves. In this way, *SnowVision* can be viewed as inspiring a type of crowd-sourced data collection. The potential for this is beautifully illustrated by several recent NEH-supported initiatives such as Scribe (http://scribeproject.github. io/). We are already seeing a high level of interest among colleagues, expressed through inquiries about contributing content to the project. Indeed, Mr. Keith, named herein as an advisory committee member, obtained a small grant from the Modern Heritage Foundation to begin to digitize the Leake site collection housed at the University of West Georgia to contribute content to *SnowVision*. Thus, we are at a critical point in the project where knowledge of *SnowVision* is spreading and interest is growing. Implementation funds are essential to making *SnowVision* fully available to researchers and the public, alike.

DATA MANAGEMENT PLAN

The development and maintenance of a comprehensive and accurate data archive is critical in meeting the objectives of this proposed project. The Project Directors have track records for prompt publication of research data and have actively shared and communicated the results with the scientific community in conferences or journal publications and via various synergistic activities.

Expected Data & Software to be Managed

The expected data and software in this project include the collected sherd and design images and other important properties for these images, the source code of the programs, experimental results, and publications. We plan to manage and make available the primary source code of the software, the results produced under this project, and the associated data that describe the experimental setup, theoretical model and data-analysis methods. For the collected image datasets, we plan to negotiate with contributing institutions to secure rights to share the digital images with professional colleagues and members of the public through the World Engraved website. The data will be prepared and published promptly in the form of online open-source downloads, peer- reviewed journal/conference articles, theses, supplementary information to published manuscripts, book chapters and other print or electronic publishing formats.

We will release the developed software for sherd-to-design matching. The software will be released under a "CC BY-NC-SA 4.0" license.1 This will (1) make the software and its source code freely available to researchers and educators; (2) permit modifications of enhanced versions for noncommercial use; and (3) allow continued development and maintenance by other individuals or groups. Existing software code that comprises *SnowVision* has already been pushed to GitHub (https://github.com/SnowVision/SnowVision1.0). The final release will occur at the end of Year Two, with at least two updated releases per year thereafter.

In Year Two, we will launch a web portal that integrates the developed software with the World Engraved website. This will allow users to upload digital images of sherds from other collections. Our computer vision algorithms will run in the background and a rank of possible matches will be provided to the user at the end of the search. As mentioned above, we plan to make such new images and patterns themselves available to other users and the general public, thus creating a data commons of sherd and pattern images. The website will also be optimized for use with mobile devices, especially smartphones. The management of new images that are collected through these web applications will be exactly the same as those initially collected data for classifier training and algorithm testing.

Data and Software Formats

The source codes of the developed software will be provided as C/C++, PYTHON code (algorithm implementations for each core research component), PHP/PYTHON (for the user's interface) and HTML with embedded Javascript (for the web applications). Image data will be in their original format when they are collected, without any compression and downsizing post-processing. Properties of interest, such as location and time information, associated with each image will be stored in pure text format. Publications will be available in print from publishers or electronically in PDF format.

Period of Retention

The Center for Digital Humanities (CDH) at University of South Carolina, led by Dr. Matthew J. Simmons, will commit to supporting the project—including security updates and maintaining the site on our GCP infrastructure—for 5 years, or the length of the grant, whichever is longer, up to 7 years. CDH is willing to continue to support the site after that time period. However, as internet and web technologies change so rapidly and regularly, predicting how long-term our support can be is impossible; a website built in 2019 may

[1] Attribution-NonCommercial-ShareAlike 4.0 International license

not function on the standard technologies of 2030. The Research Computing (RCI) group at University of South Carolina, led by Mr. Paul Sagano, will commit to supporting the project—including collaborating with XSEDE, administrating the RCI student developer and supporting additional HPC resources —for the length of the grant.

We expect to keep updating the developed software and sherd image database for a minimum of three years after the completion of the project by incorporating the reported bugs and new query images from researchers and the general public.

Data Storage and Preservation

Published data will be available in print or electronically from publishers and subject to subscription or printing charges. All software source code and test data will be stored electronically in computers installed at the offices of the PDs as well as the proposed HPC storage server. Disk space will be exclusively used for the storage of the data produced in this project. In case of changes in the roles and responsibilities of the personnel involved (such as graduation, separation from the current position), the Project Director assumes responsibility of the data.

SUSTAINABILITY PLAN

Our project will have few financial needs beyond the grant period.

As discussed in the narrative, with NEH support, we will host and manage the World Engraved website on a Google Cloud Platform (GCP) infrastructure. The flexibility of cloud-based hosting will be a boon for the sustainability of the project. Downtime/the site being offline will be minimized to industry-leading levels, thus ensuring users will, for all intents and purposes, always be able to access the site. Cloud computing also allows for flexibility in compute resources; that is, processor cores and memory can be added on-the-fly according to user demand and project needs.

We will leverage high-performance computing (HPC) facilities to speed up the large-scale matching task. In particular, we will use both the parallel job schedulers (e.g. SLURM) and machine learning packages (e.g. Caffe, Tensorflow) on GPU to improve the performance. To serve this purpose, we will request HPC resources from the Extreme Science and Engineering Discovery Environment (XSEDE), the national NSF-funded virtual system that share resources like supercomputing, data and tools. XSEDE provides a science gateway, or a set of tools that help assemble resources like HPC computing, scientific applications and data, and build automatic workflows via a portal, usually in a graphical user interface. We will request and customize a XSEDE science gateway to integrate our World Engraved website with the backend *SnowVision* matching algorithms running on XSEDE HPC clusters. A proposal requesting 1-year-startup resources to XSEDE has been approved, and now is utilized by our research team. Should our NEH proposal be approved, we will further request advanced resources at the research level from XSEDE to continue supporting our *SnowVision* project. Implementing the *SnowVision* project on a XSEDE platform can foster collaborations among researchers across domains, and most of all, be cost-effective.

Funds from the grant will be used to help pay for the project's share of website hosting at an estimated cost of $3,500 plus the cost of a student system administrator to run the website. The estimation is based on 1 1TB GCP virtual machine with 4GB memory for the course of 3 years. The Center for Digital Humanities (CDH) at University of South Carolina, led by Dr. Matthew J. Simmons, will commit to supporting the project—including security updates and maintaining the site on our GCP infrastructure—for 5 years, or the length of the grant, whichever is longer, up to 7 years. CDH is willing to continue to support the site after that time period. However, as internet and web technologies change so rapidly and regularly, predicting how long-term our support can be is impossible; a website built in 2019 may not function on the standard technologies of 2030.

Dissemination is a top priority. We believe the formation of an advisory committee of professionally invested, highly accomplished scholars early in the project is a critical step in the plan for distribution and widespread acceptance of our program among research communities. Comprised of major figures in Swift Creek pottery studies and in digital archaeology, our advisory committee will broaden the knowledge base of the team, which will serve to extend the utility of the program and encourage downstream use of it at other institutions. Also, by designing a workflow that enables much of the digitization work to be done remotely in labs and curation facilities, we distribute the costs of growing the archive beyond the data needed to fully develop and vet the program.

Technical staff at our three research centers (CEC, CDH, and RC) will monitor future development of the software. The program itself will undergo further refinement when we extend the application to sherd-to-sherd matching for purposes of the reconstruction of new designs, which is part of our longer-term goal. In this stage, we will consider requests for new features from the user community in order to enhance the value of the database.

APPENDIX A. BIBLIOGRAPHY

Berelov, I.
2006 Signs of Sedentism and mobility in an agro-pastoral community during the Levantine Middle Bronze Age: Interpreting site function and occupation strategy at Zahrat adh-Dhra‘ 1 in Jordan. *Journal of Anthropological Archaeology* 25(1):117-143.

Binford, Lewis R.
2001 *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building using Hunter-Gatherer and Environmental Data Sets*. University of California Press, Berkeley.

Broyles, Bettye J.
1968 Reconstructed Designs from Swift Creek Complicated Stamped Sherds. *Southeastern Archaeological Conference Bulletin* 8:49-55.

Chen, Liang-Chieh, et al.
2016 Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv preprint arXiv:1606.00915.

Cohen, F., E. Taslidere, Liu Zexi, and G. Muschio
2010 Virtual Reconstruction of Archaeological Vessels using Expert Priors and Surface Markings. *Computer Vision and Pattern Recognition Workshops* 2010:7-14.

Cooper, D., A. Willis, S. Andrews, and J. Baker
2001 Assembling Virtual Pots from 3D Measurements of their Fragments. *Proceedings of the Conference on Virtual Reality, Archeology, and Cultural Heritage* 2001:241-254.

Feest, Christian F.
1980 *Native Arts of North America*. Oxford University Press, New York.

Garcia-Pedrajas, N., Hervas-Martinez, C., Munoz-Perez, J.
2003 Covnet: a cooperative coevolutionary model for evolving artificial neural networks. *IEEE Transactions on Neural Networks* 14(3), 575–596.

Holmes, W. H.
1894 Caribbean Influence in the Prehistoric Art of the Southern States. *American Anthropologist* 7:71–79.

1903 Aboriginal Pottery of Eastern United States. *Bureau of American Ethnology Annual Report* 20:1-201.

Kelly, Robert L.
1992 Mobility/Stability: Concepts, Archaeological Measures, and Effects. *Annual Review of Anthropology* 21:43-66.

Knight, Vernon J.
2013 *Iconographic Method in New World Prehistory*. Cambridge University Press.

Koller, D., and M. Levoy
2006 Computer-aided Reconstruction and New Matches in the Forma Urbis Romae. *Bullettino Della Commissione Archeologica Comunale di Roma* 2006:103–125.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton
2012 Imagenet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems* 1:1097-1105.

Koutsoudis, Anestis, George Pavlidis, and Christodoulos Chamzas
2010 Detecting shape similarities in 3D pottery repositories. *IEEE Fourth International Conference on Semantic Computing* (ICSC):548-552.

Lami, Martina, Loes Opgenhaffen, and Ivan Kisjes
2015 Pottery goes digital: 3D laser scanning technology and the study of archaeological ceramics. *Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology* 2015:421-431.

Liao, Ping-Sung, Tse-Sheng Chen, and Pau-Choo Chung
2001 A fast algorithm for multilevel thresholding. *Journal of Information Science and Engineering* 17.5: 713-727.

Long, J., E. Shelhamer, and T. Darrell
2015 Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2015:3431–3440.

Lu, Yuhang, Jun Zhou, Jun Chen, Jing Wang, Karen Y. Smith, Colin Wilder, and Song Wang
2018 Curve-Structure Segmentation from Depth Maps: A CNN-based Approach and Its Application to Exploring Cultural Heritage Objects. Association for the Advancement of Artificial Intelligence (AAAI) Conference of Artificial Intelligence (in press).

Powers, David Martin
2011 Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies* 2(1):37-63.

Roosevelt, Anna Curtenius, and James G. E. Smith, Editors
1979 *The Ancestors: Native Artisans of the Americas*. Museum of the American Indian, New York.

Sassaman, Kenneth E., and Wictoria Rudolph
2001 Communities of Practice in the Early Pottery Traditions of the American Southeast. *Journal of Anthropological Research* 57:407-425.

Saunders, Rebecca
1986 Attribute Variability in Late Swift Creek Phase Ceramics from Kings Bay, Georgia. Unpublished Master's thesis, Department of Anthropology, University of Florida, Gainesville.

Shen, Wei, Kai Zhao, Yuan Jiang, Yan Wang, Zhijiang Zhang, and Xiang Bai
2016 Object skeleton extraction in natural images by fusing scale-associated deep side outputs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Shin, Hijung, Christos Doumas, Thomas Funkhouser, Szymon Rusinkiewicz, Kenneth Steiglitz, Andreas Vlachopoulos, and Tim Weyrich
2012 Analyzing and Simulating Fracture Patterns of Theran Wall Paintings. *Journal on Computing and Cultural Heritage* Vol. 5(3), October 2012, Article No. 10.

Smith, Karen Y., and Vernon J. Knight, Jr.
2014 Core elements and layout classes in Swift Creek paddle art. *Southeastern Archaeology* 33(1):42-54.

2017 Swift Creek paddle designs and the imperative to be unique. *Southeastern Archaeology* 36(2):122-130.

Smith, Karen Y., and Keith Stephenson
2018 The spatial dimension of the woodland period. *Southeastern Archeology* 37(2):112-128.

Snow, Frankie
1975 Swift Creek Designs and Distributions: A South Georgia Study. *Early Georgia* 3:38-59.

1977 An Archeological Survey of the Ocmulgee Big Bend Region. *Occasional Papers from South Georgia*, Number 3. South Georgia College, Douglas.

1982 Kolomoki and Milamo: Some Synchronic Evidence. *The Profile* 36:7-12.

1998 Swift Creek Design Investigations: The Hartford Case. In *A World Engraved: Archaeology of the Swift Creek Culture*, edited by Mark Williams and Daniel T. Elliott, pp. 61-98. University of Alabama Press, Tuscaloosa.

2003 Original Unpublished Swift Creek Design Catalog. Manuscript on File, South Georgia State College, Douglas, GA.

Snow, Frankie, and Keith Stephenson
1998 Swift Creek Designs: A Tool for Monitoring Interaction. In *A World Engraved: Archaeology of the Swift Creek Culture*, edited by Mark Williams and Daniel T. Elliott, pp. 99-111. University of Alabama Press, Tuscaloosa.

Stephenson, Keith, Judith A. Bense, and Frankie Snow
2002 Aspects of Deptford and Swift Creek of the South Atlantic and Gulf Coastal Plains. In *The Woodland Southeast*, edited by David G. Anderson and Robert C. Mainfort, Jr., pp. 318-351.

Tehrani, J. and F. Riede
2008 Towards an archaeology of pedagogy: learning, teaching and the generation of material culture traditions. *World Archaeology* 40(3):316-331.

Wallis, Neill J.
2011 *The Swift Creek Gift: Vessel Exchange on the Atlantic Coast*. University of Alabama Press, Tuscaloosa.

2013 The Materiality of Signs: Enchainment and Animacy in Woodland Southeastern North American Pottery. *American Antiquity* 78(2):207–226.

Wallis, Neill J., and Amanda O'Dell
2011 Swift Creek Paddle Designs from the Florida Gulf Coast: Patterns and Prospects. *The Florida Anthropologist* 64(3-4):187-205.

Wauchope, Robert
1966 *Archaeological Survey of Northern Georgia, with a Test of Some Cultural Hypotheses*. Number 21. Memoirs of the Society for American Archaeology.

Wilson, Eva
1984 *North American Indian Designs*. British Museum Publications, London.

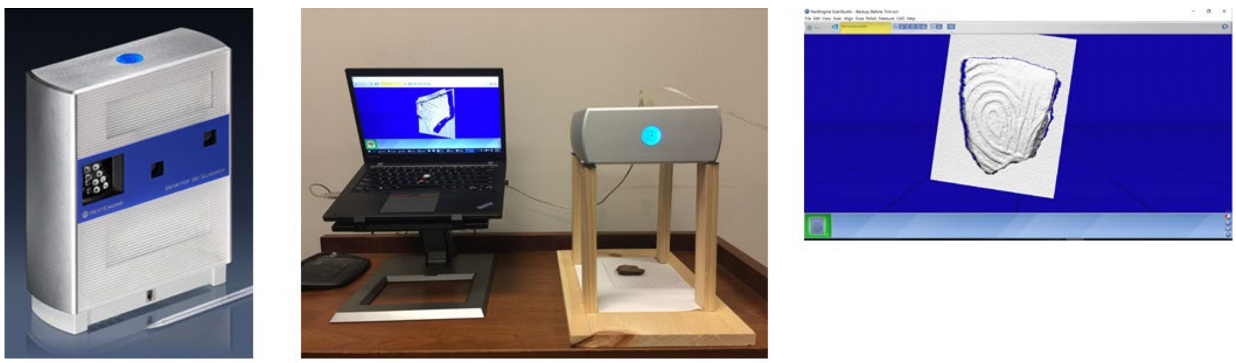
Zhou, Jun, Haozhou Yu, Karen Y. Smith, Colin Wilder, and Song Wang
2017 Identifying Designs from Incomplete, Fragmented Cultural Heritage Objects by Curve-Pattern Matching. *Journal of Electronic Imaging* 26(1). DOI10.1117/1.JEI.26.1.011022.

Vese, Luminita A., and Tony F. Chan
2002 A multiphase level set framework for image segmentation using the Mumford and Shah model. *International journal of computer vision* 50.3: 271-293.

Vidal, Aixa
2017 Which way? Handedness in ceramic decoration. In *Playing with Time: Experimental Archaeology and the Study of the Past*, edited by Rodrigo Alonso, Javier Baena & David Canales, pp. 261-268. Servicio de Publicaciones de la Universidad Autónoma de Madrid. Madrid.

APPENDIX B. 3D POTTERY SHERD DIGITIZATION

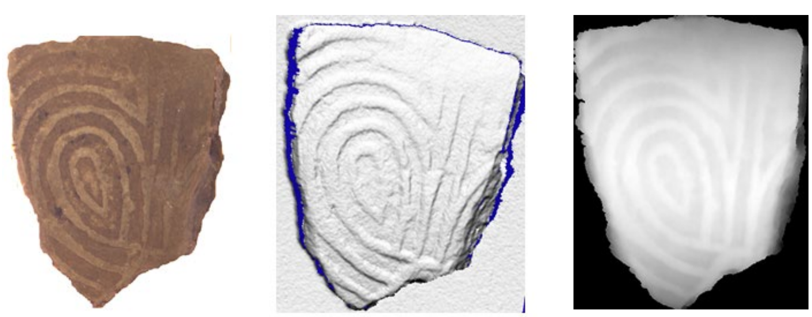
For this project, we have developed a digitization protocol to scan pottery sherds surfaces. We are using a low-cost, efficient scanner, the NextEngine 3D Ultra HD Laser Scanner (shown in Figure 1a) and its included ScanStudio application (shown in Figure 1c), which manages the scanner hardware, refines the output scanned data, and assembles the data into a healed 3D mesh model, e.g. .object point cloud. ScanStudio is installed on a Windows laptop of 2.5 Quad Core, 16GB RAM and 120GB hard drive. The scanner is placed 9 inches above the sherd and perpendicular to the sherd platform, illustrated in Figure 1b. We scan the sherd's surface with the resolution of 100 points per mm2.



(a) (b) (c)

Figure 1. (a) NextEngine 3D Laser Scanner. (b) Scanning Setup. (c) ScanStudio Application with an example.

Furthermore, we have developed a program to generate pottery sherd depth map image based on the 3D scanned output file. The depth map contains information relating to the distance of a pottery sherd surface from the center of the NextEngine Scanner. We take the 3D file (in widely-used .obj format) and sample it with the same resolution as in scanning process, i.e., each pixel in depth image covers 0:01mm2. This program was implemented using Visual C++ and Point Cloud Library (PCL). Figure 2 shows a sherd sample and its generated depth map image.



(a) (b) (c)

Figure 2. (a) A pottery sherd RGB image. (b) The 3D point cloud image. (c) The depth map image.

APPENDIX C. WEBSITE WIREFRAMES

The website (www.worldengraved.org, named after a scholarly volume on Swift Creek) features a way for registered users to submit sherd data and images (Figure 1) and a way for users to query and download existing sherd and design data (Figure 2).

Figure 1. Users do not need to be logged in to query the image and design database.

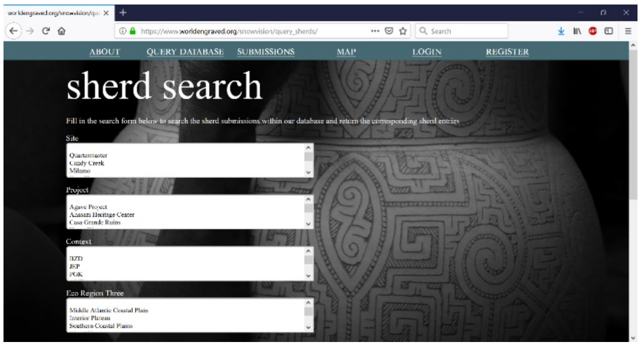


Figure 2. Registered users can query and download sherd metadata and images. Queries return metadata and provide access to a details page.

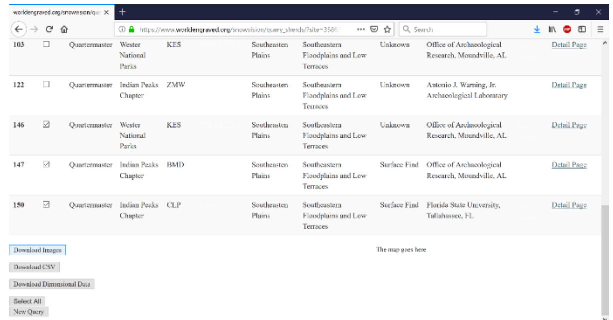
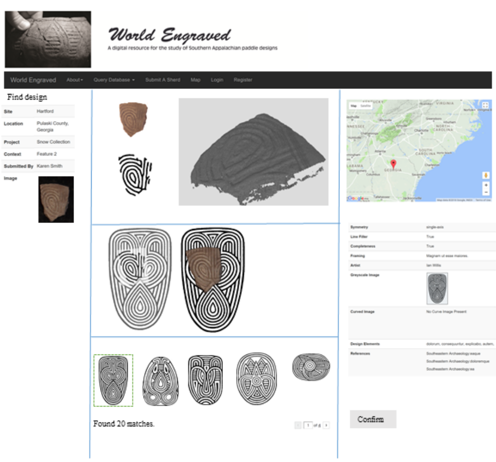
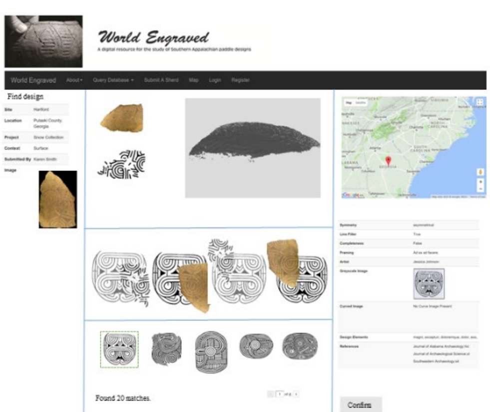


Figure 3. Hypothetical results returned for a single, non-composite sherd.



On the left is a draft of the way design matches will be returned to users. Figure 3 concerns sherds with a single paddle impression pattern. Figure 4 concerns sherds with composite, or overstamped, impressions.

Figure 4. Hypothetical results returned for a composite sherd.



APPENDIX D. CURVE STRUCTURE SEGMENTATION

New forms of computation such as artificial intelligence, particularly machine learning have started changing historical notions of humanities research and thinking more and more. *SnowVision* is such a tool. It is an integration of deep learning – an advanced machine learning technique, and traditional archeological heritage fragment recognition and reconstruction.

SnowVision studies cultural heritage fragments and their corresponding designs, especially pottery sherds excavated in southeastern North America and the Native American paddle designs that are partially presented on these sherds's surfaces. *SnowVision* automatically identifies the underlying paddle designs of these pottery sherds from a database of known designs. Usually, in the making of pottery, Native Americans applied the paddles on pottery surfaces multiple times. As a result, the pottery sherds many bear multiple overlapping patterns, each is a part of design pattern as shown in Figure 1(3). We define a sherd with a single design pattern as non-composite pattern sherd shown in Figure 1(2)(a), and one with multiple overlapping patterns as composite pattern sherd shown in Figure 1(3)(a). Illustrated in Figure 1, the identification process consists of three parts: curve structure segmentation, non-composite curve pattern matching and composite curve pattern matching.

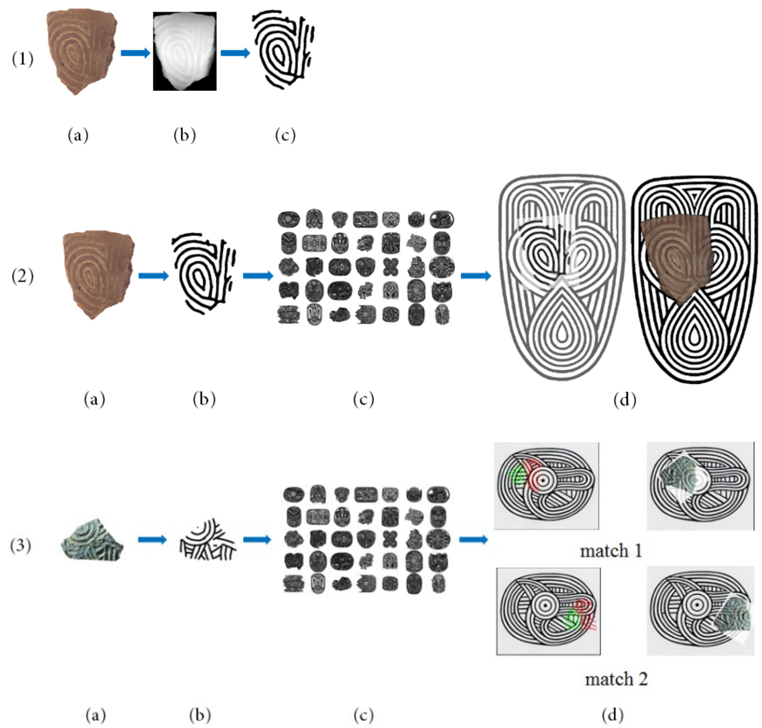


Figure 1. Underlying design identification. (1) Curve structure segmentation: (a) a sherd RGB image, (b) the depth map image of the sherd, and (c) curve structure extracted; (2) Non-composite pattern matching: (a) a sherd with non-composite pattern, (b) the non-composite curve pattern extracted, (c) design database, and (d) matching result; (3) Composite pattern matching: (a) a sherd with composite pattern, (b) the composite curve pattern extracted, (c) design database, and (d) matching result.

The following section describes the developed new technical method for curve structure segmentation. Our specific, computational approach to curve pattern matching is detailed in a paper published by the Journal of Electronic Imaging (Zhou et al. 2017).

Curve Structure Segmentation

Accurately segmenting curve structures stamped on pottery surfaces is the first step to explore cultural heritage fragments. In most cases, these curve structures do not bear distinctive colors, and it is very difficult, if not impossible, to segment them from an RGB image of a sherd. Thus we use a 3D scanner to produce a depth map image of the sherd surface. The locations of curves exhibit a larger depth than the non-curve portion of surface. However, segmenting curve structure from the depth map image is very challenging because the curves can be very shallow or contain a lot noise due to sherd erosion or highly-rough pottery surfaces. As the result, it is very difficult to segment these curve structures using existing low-level methods.

For this project, we have developed a new deep-learning algorithm to segment such curve structures that were weakly stamped on the pottery surfaces. Specifically, it contains three steps, as illustrated in Figure 2. In the first step, a fully convolutional neural network (FCN) (Krizhevsky et al. 2012) is employed to extract the skeleton of curve structures, and estimate a scale value at each skeleton pixel. This scale value reflects the curve width at the corresponding skeleton pixel. In the second step, we propose a dense prediction network to refine the curve skeletons. In the third step, we develop an adaptive thresholding algorithm (Liao et al. 2001) to achieve the final segmentation of curve structures with width by considering the estimated scale values.

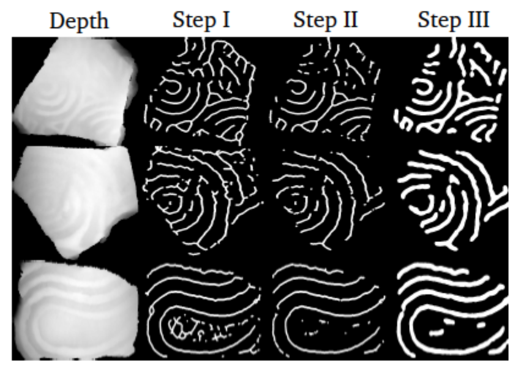


Figure 2. Three sample sherds: depth map images and their curve structure segmentation results after each step.

In Step I, a FCN is designed and trained to extract the skeleton of curve structures, illustrated in Figure 3. In depth map image, we label skeleton pixels as 1 and non-skeleton pixels as 0. Then curve structure segmentation is the same as classification using the labels. The FCN contains 3 encoders and 3 decoders (Long et al. 2015). Each decoder follows one encoder. Each encoder is a small convnet (Garcia-Pedrajas et al. 2003) containing different size of convolutional layers, ReLU layers and max-pooling layers. The encoder is applied to extract features for further classification by its decoder. The use of multiple encoders/decoders is to extract features at different levels. The output of each decoder is then used to generate skeleton heat map S, and a skeleton map is produced by thinning the heat map.

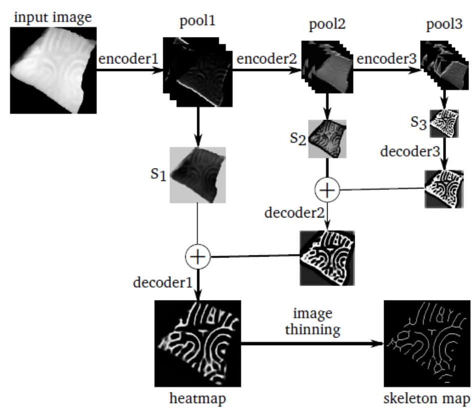


Figure 3. The FCN used to extract skeleton map.

In Step II, we refine the curve skeleton using a dense prediction convnet. The skeleton detected in Step I may still contain false positives (noises). To prune these false positives, we train a soft classifier where a skeleton probability is outputted at each pixel. The probability map D is produced by the following transformation:

$$D(x,y) = \frac{1}{1 + \min\limits_{(x',y')\in P} \sqrt{(x-x')^2 + (y-y')}} \qquad (1)$$

, where P is the set of skeleton pixels in the binary skeleton map. This classifier is a convnet consists of three convolutional layers, three max-pooling layers and two fully connected layers. In testing, we prune the low-probability skeleton pixels to achieve a refined skeleton.

Algorithm 1 Curve-Structure Segmentation by Recovering Curve width

Input: Depth image I; Refined skeleton P, Scale values s
Output: Binary segmentation map C

Initialize all the elements in C to zero.
for each skeleton pixel $(x,y) \in P$ **do**
  Compute neighborhood:
  $\mathbb{N} = \left\{(x',y') \middle| \sqrt{(x-x')^2 + (y-y')^2} \le 2^{s(x,y)}\right\}$.
  for each pixel $(x',y') \in \mathbb{N}$ **do**
    if $I(x',y') \ge \frac{I(x,y)+\min_{(x'',y'')\in\mathbb{N}} I(x'',y'')}{2}$ **then**
      $C(x,y) = 1$
    end if
  end for
end for

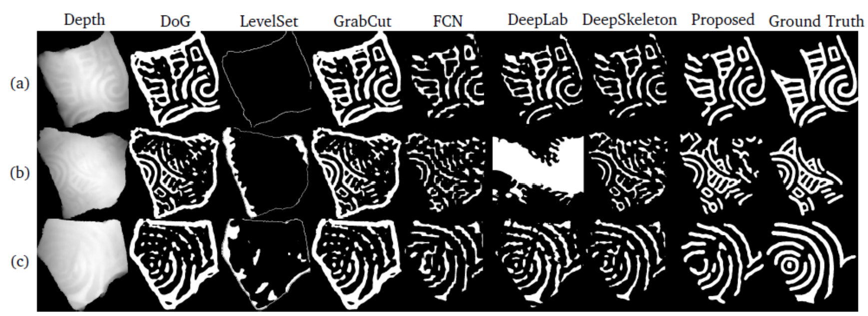
In Step III, we recover curve width of curve structure using the skeleton map derived in Step II, with the help of the scale values derived in Step I. Denote the original depth image by I and let P be the set of refined skeleton pixels detected on I after Step II, we construct the curve segmentation, in the form of a binary map C of the same size as I, using the following algorithm 1.

To test our algorithm, we have collected 1174 depth map images of pottery sherds excavated in multiple sites in Southeastern North America. We evaluated the effective of the algorithm by comparing it with six widely-used segmentation methods for comparison – Difference of Gaussian (DoG) (Lindeberg 2015), Level Set (Vese et al. 2002), GrabCut (Rother et al. 2004), Fully Connected Network (FCN) (Long et al. 2015), Deep Skeleton (Shen et al. 2016) and DeepLab (Chen et al. 2016). Figure 4 shows the comparison results of three sample sherds.

The table (right) shows the proposed algorithm and six comparison methods in terms of F-measure (Powers et al. 2011). As we can see, our new method outperforms existing methods significantly. The curve structure segmentation method and its experiment results are detailed in our paper published in the Association for the Advancement of Artificial Intelligence (AAAI) 2018 (Lu et al. 2018).

Methods	Precision	Recall	F-measure
DoG	0.423	0.811	0.547
LevelSet	0.065	0.085	0.066
GrabCut	0.402	0.711	0.517
FCN	0.589	0.450	0.503
DeepLab	0.585	0.670	0.581
DeepSkeleton	0.687	0.598	0.627
Proposed	0.670	0.786	**0.717**

Figure 4. Examples of the curve-structure segmentation result from the proposed method and six comparison methods.



APPENDIX E. EXAMPLES FROM ADJACENT FIELDS THAT COULD USE OUR ALGORITHM

Identifying symbols stamped on ancient coins. For example, shown below: from left to right: (1) the symbol, (2)-(5) top: ancient coins, down: matching result. Symbols are marked in light red; matching result: (2) and (3) find their correct locations, (4) close to its correct location , and (5) not find the symbol.

(1) (2) (3) (4) (5)

Comparing multiple copies of Doctrina Christiana wood prints. (1) the source copy of the print, (2)-(4) left: the target copy of prints, right: the source copy overlaying the target copy. Note that our algorithm first finds its correct location and then overlays these two copies. The blurring and missing lines indicate the difference between copies.

(1) (2) (4) (3)

APPENDIX F. EVALUATION SURVEY (DRAFT)

The following short questionnaire will be emailed to members of the advisory committee and to other potential users during the final months of the grant.

1. Have you heard of *SnowVision*? If yes, how (through what medium)?

2. Have you visited www.worldengraved.org? If yes, how often? Once a week, once a month, once a quarter, only once, other.

3. If you have visited the site, do you find it easy to navigate and use?

4. How likely are you to use the website in the future for research or teaching?

5. How likely are you to use the *SnowVision* matching component of the website now or in the future?

6. How likely are you to contribute content to the *SnowVision* database in the future?

7. For current users, what improvements to either the website or *SnowVision* would you like to see?

8. What additional tools or resources would enable you to contribute more content through time?