Towards Standards for the Manual Digitization and Vectorization of Analog Historical Maps and Records

Ian Dunshee
University of Iowa

Follow this and additional works at: https://ir.uiowa.edu/honors_theses
Part of the Archaeological Anthropology Commons, and the Geographic Information Sciences Commons

Copyright © 2016 Ian Dunshee

Hosted by Iowa Research Online. For more information please contact: lib-ir@uiowa.edu.
TOWARDS STANDARDS FOR THE MANUAL DIGITIZATION AND VECTORIZATION OF ANALOG HISTORICAL MAPS AND RECORDS

by

Ian Dunshee

A thesis submitted in partial fulfillment of the requirements for graduation with Honors in the Anthropology

__________________________________________________________________________

James G. Enloe
Thesis Mentor

Summer 2016

All requirements for graduation with Honors in the Anthropology have been completed.

__________________________________________________________________________

Margaret Beck
Anthropology Honors Advisor

This honors thesis is available at Iowa Research Online: https://ir.uiowa.edu/honors_theses/
TOWARDS STANDARDS FOR THE MANUAL DIGITIZATION AND VECTORIZATION
OF ANALOG HISTORICAL MAPS AND RECORDS

By
Ian Cole Dunshee

A thesis submitted in partial fulfillment of the requirements for graduation with Honors in the
University of Iowa, Department of Anthropology

Dr. James G. Enloe
Thesis Mentor

Summer 2016
All requirements for graduation with Honors in the Department of Anthropology have been completed.

[Signature]

Dr. Margaret Beck
Department of Anthropology Honors Advisor

College of Liberal Arts and Sciences Honors Director
Introduction

Archaeology is an important field which studies human history and prehistory through the excavation and analysis of physical remains as they exist in the present (Chazan 2014). This requires careful attention to and recording of not only the condition and material of an object found during an archaeological excavation but its spatial context (location and orientation) as well. By evaluating the integrity of an archaeological deposit (Schiffer 1973) the spatial information contained therein can be interpreted to provide information regarding human activities and life in the past including adaptive behaviors, cultural and social practices, and even evolutionary changes (Rapson 1990).

In the past 40 years, GIS (Geographic Information Science) and Remote Sensing techniques (most notably LiDAR, a 3D environment scanning technology, and Ground Penetrating Radar) have proved to be effective tools in the pursuit of a more accurate and precise archaeological practice due to their ability to record, digitize, analyze, and visualize geo-spatial information and preserve archaeological context in ways not possible before (Wiseman and El-Baz 2007). With archaeological data in a digital format, computer systems can be used to perform and create more complex, complete, subtle, and large scale analyses and visualizations that were not available in previous years and have provided a great increase in the information gained from a site.

Despite this, archaeologists have been required to utilize and manipulate software tools designed for other tasks in order to achieve their goals because very few have been written specifically with archaeological needs in mind. This has created a situation in which archaeologists have needed to shift their focus to learn basic GIS techniques while many skilled Geographic Information and Remote Sensing specialists have little interest in the needs of archaeologists. One such area in which this is particularly evident is in the digitization and vectorization of analog archaeological records.

Digitization converts analog maps to a digital format using a scanner, digital camera, or other similar device (Esri “Digitizing”). The digital format they are converted to is known as a raster, “An abstraction of the real world where spatial data is expressed as a matrix of cells or pixels, with spatial position implicit in the ordering of the pixels. The pixel value indicates the attribute such as color, elevation, or ID number,” (Dharmaraj 2005). Rasters can be converted into vector data, another leading standard format for spatial data which stores information in the form of points, lines, or polygons bounded by coordinates and to which attributes are given (Dharmaraj 2005) This conversion process is commonly referred to as vectorization and can be done automatically or manually on a heads-up display or computer screen (Esri “Vectorization”). There are still many historical and archaeological records that would benefit from the computer-based analyses that could be performed as a result of this conversion because of the profession’s long history prior to the development of computer technology as we know it today. However many of these maps cannot be viewed or analyzed with the benefit of this technology because of a lack of effective algorithms or standards for their conversion.
Algorithms, or, ”A procedure for solving a mathematical problem in a finite number of steps… a step-by-step method for accomplishing some task,” can be used to create computer software and programs which can automate processes like digitization and vectorization. When manual methods for a regularly needed task prove too time-consuming and costly to pursue, an algorithm can be created, but this too requires a great deal of time and effort and so is not undertaken for everything (Schneider and Gersting 2015). In regards to the digitization and vectorization of historic archaeological maps and others like it, little attention is given to the creation of the more complex algorithms they require and so the only option is a manual method. In cases in which manual digitization and vectorization is undertaken, the lack of universally accepted methods and standards can make the error and results of such projects difficult to compare (Skalos et al. 2011; Nobajas and Nadal 2015; Kaim et al. 2016; Daniels and Huxford 2001; Schuppert and Dix 2009).

Though many tools and algorithms exist for the vectorization of analog maps, they all are designed to handle specific complexities that differ from those contained in most analog historical and archaeological records. For instance, many current algorithms exploit the use of continuous lines (as in topographical maps) and map colors, both of which do not frequently occur in monotone archaeological maps and sketches containing small, irregular, and discontinuous features. Even when used for the tasks they were designed for, the current programs and algorithms often require manual intervention to check for and correct errors as well as to vectorize areas and features not detected by the algorithm itself (Sharma 2006; Chiang 2010; Mabee et al. 2003; Kaim et al. 2016; Wu 1998; Dharmaraj 2005; Daryal 2010; Oka et al. 2012; Yang et al. 2012; Walls 2015). Because of this, the current automatic methods of vectorization are less useful to archaeologists.

This means that manual vectorization is still the preferred method when dealing with historical and archaeological analog maps. This presents a unique set of problems to be dealt with because they are much more irregular than modern standardized maps (Nobajas and Nadal 2015; Skalos et al. 2011; Kaim et al. 2016; Daniels and Huxford 2001; Schuppert and Dix 2009). However, in the current literature, very little description if any is given to the specifics of the manual digitization or vectorization methods used in recent studies (Nobajas and Nadal 2015; Skalos et al. 2011; Kaim et al. 2016; Daniels and Huxford 2001) and there are no apparent or proposed standards.

In this study, a set of easily replicable and clear standards for the manual digitization and vectorization of analog; historical and archaeological maps is proposed. The method is applied using records from the Woodpecker Cave archaeological site (13JH202) made during the 2012-2015 field seasons as examples. The result of the methodology supplements and adds to the current site records and contributes toward the broader goal of using spatial organization of remains at the site to gain a better understanding of how seasonal sites played into human ways of life in the late Woodland period in Iowa. In addition, an accuracy assessment of the method as well as the records themselves is made and is followed by a brief discussion of the benefits and further use and development of the techniques presented herein. These techniques are not
intended to cover all possibilities and situations but are presented with the hope of future development and refinement in mind.

**Background**

Woodpecker cave (13JH202) is a Mid to Late Woodland period rockshelter site located in Coralville, Iowa. Excavated previously by Warren Caldwell (1961), the site is currently excavated by the University of Iowa Archaeological Field School (Enloe 2015, 2012-2015a/b). The results up to the present from both teams of excavators suggest a seasonal occupation primarily in Fall and Winter to procure game resources when horticulture and agriculture was not viable. This conclusion has been reached in part due to the spatial organization of artifacts at the site which show distinct and separate areas for various activities such as cooking, skinning, and tool production. This differs from the more continuous deposits that would be seen at a site of more prolonged habitation.

The spatial records kept for the excavations include both hand-drawn and printed maps though the maps from the Caldwell excavations were not made for or converted to a form useable in a Geographic Information System because Geographic Information Systems did not then exist in the form they do today. The artifact distribution images from the University of Iowa Field School, an example of which can be seen in Figure 24, only include the artifacts visible at the end of each excavation depth level reached and so do not show all artifacts found in a unit. In addition, the images were created in image processing software such as Adobe Photoshop (Adobe Systems Incorporated 2016) and so do not contain spatial information that could be immediately exploited and used by GIS software. For these reasons, the field records of the site had a need to be converted into a form that would not only show all artifacts but also be useable in GIS software through common methods such as digitization and vectorization.

Digitization and vectorization are key elements in a wide variety of GIS projects because it allows for a great increase in potential for analysis and visualization. By transforming data from a physical form to a digital form, the information can be manipulated and analyzed in a more iterative, complex, and repeatable/representative way that differs from what is possible by an individual person. This is true for many areas of study ranging from topographical maps, historical and social projects, road maps, urban planning, and much more.

Generally, the process of digitization, vectorization, and analysis in studies such as this one follow similar steps. The first step is obtaining the raster image from the analog via a digital photograph or a document scanner. The resolution and quality of a raster is determined by the resolution of the recording device as well as the quality or resolution of the data source. The resolution of the raster image can have a strong impact upon its geometric accuracy and thus the accuracy of anything derived from it. (Dharmaraj 2005). This is also the case for any warping or distortion caused by a scanner malfunction or irregular practice in the physical placement of a document while scanning (such as movement during a scan).

It is for these reasons that a source and scanning technique of the highest resolution available is ideal for projects that require high-fidelity to their sources like those involving historic archaeological documents and others like it because the context and information which
they represent is destroyed in the excavation process. It is important to realize that the initial step of digitization is important in preventing “error propagation” or the persistence and accumulation of error throughout a series of operations and outputs beginning from an original input dataset (Esri “Error Propagation”). The lower the resolution of a digitized source, the lower the resolution (and accuracy) of any resulting data from operating on that digitized source. Across multiple operations with their own potential additions to the error, the error could increase to a level that could affect the accuracy and usefulness of the results.

The second step involves manually georeferencing the raster image using appropriate software such as ArcMap (Esri 2015), Erdas Imagine (Hexagon A.B. 2015), or other similar systems designed to deal with spatial data. Georeferencing generally is the process of, “Aligning geographic data to a known coordinate system so it can be viewed, queried, and analyzed with other geographic data… [it] may involve shifting, rotating, scaling, skewing, and in some cases warping, rubber sheeting, or orthorectifying the data,” (Esri “Georeferencing”). This can be done with any spatial data type including raster and vector data though it is often confused with georectification in which two raster datasets are aligned using multiple control points (Esri “Georectification”). This is primarily because a georectification type process can be used to accomplish georeferencing tasks and such a method can be particularly effective for aligning datasets that include variety of file types.

When the needed datasets are aligned, the vectorization or tracing of the georeferenced image can take place either manually or automatically with algorithms and computer programs, both of which include the manual correction of any errors that occur (Dharmaraj 2005; Schuppert and Dix 2009). In basic terms then, the entire process involves scanning of the map to be vectorized, alignment of this map with known geographic coordinates via georeferencing or a similar process, vectorization of the map elements once georeferenced, and any error correction as needed. After this, the digitized and vectorized map can be used for more complex analyses in a Geographic Information System. While the scanning, correction of errors, and georectification/georeferencing are generally straightforward and most commonly done manually, the benefits and challenges of manual vs. automatic methods of vectorization is a vibrant area of study today (Sharma 2006; Mabee et al. 2003; Daniels and Huxford 2001; Dharmaraj 2005; Balaji and Punithavalli 2010; Daryal 2010; Oka et al. 2012; Yang et al. 2012).

However, due in part to the difficulty of performing an accuracy assessment on manually vectorized data when no vectors or “true” locations to compare them with exist and due to the consensus that manual methods are a less accurate and imprecise alternative, the most recent studies tend to focus on the role of automatic vectorization (Chiang 2010; Sharma 2006; Mabee et al. 2003; Dharmaraj 2005). Through the continual improvement of automatic vectorization algorithms this may indeed prove true, however currently such automatic methods are still in development and while in some cases they only require correction, in others they are completely unusable. This is because the complex nature of their work can only be used with datasets of a particular type and still contain many errors which need to be manually corrected which eliminates any benefit they could offer in saving vectorization time if they can be used at all.
For example, many such automatic vectorization algorithms are designed to exploit continuous road or topographical lines (Dharmaraj 2005) as well as map color/morphology of colored parts (Sharma 2006; Mabee et al. 2003; Kaim et al. 2016; Oka et al. 2012; Yang et al. 2012) or greyscale (Chiang 2010). This means that in any situations which deviate from these conditions, the automatic methods are not of much use including cases in which maps are colorless or contain irregular and discontinuous objects (like many historical and archaeological maps). The current algorithms also make extensive use of techniques such as line thinning and approximation which exploit and change the map morphology resulting in error (Daryal 2010) which in some cases needs to be corrected (Sharma 2006; Chiang 2010; Mabee et al. 2003; Kaim et al. 2016; Wu 1998; Dharmaraj 2005; Daryal 2010; Oka et al. 2012; Yang et al. 2012). While this error is at an acceptable level for the large scale topographical and road maps they are concerned with, they are not as suitable for archaeological endeavors for which retention of exact information is paramount as the context of the sites depicted in archaeological maps are all, by virtue of being excavated, destroyed. Finally, the abovementioned studies as well as that of Walls (2015) makes clear that use of automatic vectorization techniques always requires manual intervention for the correction of errors; a task as time consuming and laborious as manual vectorization.

It is not only because of the input limitations and errors of automatic vectorization that some projects turn to manual methods. Automatic vectorization requires many extra steps to simplify or break down the problem for the algorithm/computer. For example, issues of separation of text and map lines (Chiang 2010) and separation of feature layers (Chiang 2010) are time-consuming steps which are often glossed over in recent studies despite the fact that the majority of maps to be converted to vector information will contain both text and more than one feature type.

Because of the lack of universality and continuing development of automatic vectorization methods, many studies involving historic maps do turn to manual vectorization. Usually historical maps are either too complex (containing hand drawn and often very faded symbols, features, and outdated measurement systems: Nobajas and Nadal 2015; Skalos et al. 2011; Kaim et al. 2016; Daniels and Huxford 2001; Schuppert 2009) for any current algorithms or programs to account for or they are too simple, containing no predictable morphology (Enloe 2012-2015a) or definitive color or greyscale (Enloe 2012-2015a; Caldwell 1961) to exploit. This is often the case with many hand-drawn archaeological artifact/unit maps because they are created in the field and have a more technical than artistic or aesthetic purpose. In addition, many historic maps contain irregular and discontinuous shapes and lines which would pose a challenge to the current automatic methods and compromise their accuracy.

The decision to choose the manual method is also supported by the fact that automatic vectorization methods do not yet provide a valuable increase in accuracy. Wu’s (1998) accuracy assessment of an automatic vectorization technique and Daniels and Huxford’s (2001) assessment of a heads-up manual vectorization project reveal that both methods allow for error within the acceptable range of the project to produce a usable result. This then means that in their current state manual and automatic vectorization methods are generally equal in regards to
time, effort, and accuracy if an automated algorithm exists for a datatype. While automatic vectorization operates without user input, it has only a limited types of data it can work with and requires a manual check and correction process. Manual methods on the other hand, require complete user input and are subject to human error though the extra attention they require make them less prone to blatant and fatal mistakes and allow them to be used with virtually any type of map and data.

In all of the above mentioned studies which use manual digitization and/or vectorization methods, none detail the methods or standards of the process (Nobajas and Nadal 2015; Skalos et al. 2011; Kaim et al. 2016; Daniels and Huxford 2001). Because archaeological context is destroyed during excavation, it is the duty of the archaeologist to record the context of the site to its fullest extent and glean as much information as possible from those records. Establishing simple, accurate, and repeatable standards for manual digitization and vectorization is a step toward that goal. The following study aims to both establish and verify the accuracy of an easy to follow workflow for the manual digitization and vectorization of archaeological and other historic maps to enable archaeologists and others to better evaluate the integrity of archaeological deposits. This enables a more accurate and complete platform from which to interpret spatial organization at a site, which ultimately provides insight into human activity.

**Methodology**

The data used in this study included:

- A vector polygon shapefile (a digital standard of vector GIS data) of all 1x1 meter excavation units at the Woodpecker Cave field site (McGrath 2015a)

- A vector point shapefile of all artifact positions recorded with a theodolite in the field (McGrath 2015b)

- The field dossiers from field seasons 2012-2015 digitized (scanned) at 200 dpi (dots per inch) which contained the hand drawn maps created by the excavators (Enloe 2012-2015a) of which 102 were used.

All of the spatial data in the above files was projected in a non-standard locally created coordinate system of the Woodpecker Cave (WPC) field site which was based upon a local datum. All measures were done in the metric system using either meters, centimeters, or millimeters.

Firstly all scanned dossiers were examined and a list made of which units and levels of depth contained identifiable artifacts which could be traced during vectorization. This also required that each artifact to be vectorized would need to be able to match with a point file that would act as its “true” location as an error assessment. This resulted in 18 of a total of 31 units (102 total maps) spanning a maximum depth of approximately 80 centimeters below the surface (the goal being for each level of depth to be 10cm) containing a total of 338 artifacts that could be digitized for vectorization for this study. Not all of the artifacts for the site could be used because of various reasons including not all artifacts being drawn into the dossier maps by excavators, not all artifacts having identifiable labels or markings, as well as the evolution of
standards and protocols in the recording of artifacts across field seasons. It is also important to note that the maps often contained drawings of the “éboulis” or detached pieces from the overhanging rock shelter at the site though these were not included in this study to reduce vectorization time and to keep the study’s focus consistent.

Referencing the list of units and depth levels containing the artifacts to be vectorized, the hand drawn maps were cut out from the scanned field dossiers as closely as possible to their border to prevent overlap resulting from the georeferencing process that could affect later vectorization with careful note being taken of any artifact identification numbers that were cut off in the process. As can be seen in Figure 1 this was not always straightforward as some of the scans were misaligned, warped, or contained artifact shapes beyond their borders and thus the same cutout frame could not be applied to all maps. Each cutout was exported to a bitmap (.bmp) file type and converted to the Erdas Imagine Image (.img; Hexagon A.B. 2015) file type in preparation for the georeferencing process which was required to endow the scanned maps with spatial information. For this, a georectification-type method using control points was chosen because it was deemed the best way to deal with the warping of some of the maps. If not corrected, the warping would greatly affect all processes and analyses done thereafter and so alternative methods such as rubbersheeting would not be appropriate.

Figure 1
a.
Figure 1: a. a more or less aligned map with no protrusions on any side. b. a highly warped map resulting from the scanning process. c. an aligned map with an artifact protruding from its side. Only protrusions such as that seen in map c. would cause overlap in the final product but the common warping seen in map b. meant that georeferencing would need to be done manually as each map was warped at a slightly different angle. The black outlines are the map borders while the blue outlines are a straight border for comparison.
Each unit image was individually georectified in Erdas Imagine (Hexagon A.B. 2015) due to the irregularity in their borders. It is important to note that other software such as Esri’s ArcMap (2015) would work just as well for this task and the process and technique would remain the same regardless of the software used. Because the unit images were to be aligned with the WPC excavation grid shapefile depicted in Figure 2, the corners of each excavation unit from this file were used as reference points. Since the same grid shapefile would be used as a reference for all unit rectifications, the ground control points for each unit in the shapefile could be reused throughout the rectification process.

Figure 2

This was not the case for the hand drawn scanned unit maps which proved to be highly irregular in their corners and borders. This issue is the first concern which this study attempts to provide a standard practice for. The irregularity of the hand drawn unit maps seen in Figure 1 resulted in irregular raster images when scanned, all of which contained a wide variety of pixel organization along both the sides and corners of the border. With the highest level of accuracy and precision in mind it was important to determine a regular and repeatable method of accurate ground control point placement so that the georeferenced result would not contain any inaccuracies that could result in inaccuracies during the vectorization of artifacts.

Though the scanned maps appear as bitmaps, they are actually scanned images. While the precision and accuracy of later processes will likely increase with an increase in scan resolution, it is also important to think about efficiency when scanning a large number of maps as was the case here. The lowest tolerable resolution (200dpi) was chosen as a balance between accuracy needs and file size/processing time. As can be seen in Figure 3, a higher resolution does refine corners and edges, reduce protrusions, and generally make a smoother image. However, the difference between them is rather minimal and in all cases, protrusions, obscure corners, warping, and irregular lines still occur. Thus when dealing with the vectorization of digitized maps, the same issues occur at all resolutions. This is not to say that one resolution is better but...
rather that a different resolution presents a different set of problems for each dataset and anyone conducting a digitization or vectorization project needs to be critical about what problems, levels of accuracy, tolerable error, file size, processing time, manual labor time, and overall effort best suits a particular project because all of these are determined by the resolution of the original digitization.

**Figure 3**

Firstly, the corners of the excavation units were the only landmark feasible for the rectification, special attention was given to determining how to place the ground control points for each irregular corner type encountered. Examples of these are displayed in **Figure 4** which shows how ambiguous the definition of a pixelated “corner” can be.
However, this was not often the case and using a direct approach of assigning corner points to the “furthest black edge” became problematic due to various factors including but also extending beyond the shapes of the corners themselves. The “furthest black edge” approach is also problematic because a black pixel does not necessarily represent the actual end of a marking just as a white pixel does not represent the complete absence of one. Different scanners use different algorithms for deciding the value of a pixel. This can be further affected by file format and can include resampling as well as “lossy” or “lossless” techniques. However since many commercially available scanners determine a white or black pixel color based upon the majority found within a pixel it can be assumed in some cases, particularly those involving skewed and warped scans, that the “true” line lies somewhere in between the black and white pixels of the edges and corners. Because of this, other cues from the scanned image and original images (if available) also need to be taken into account.

Figure 4: As can be seen here in many cases a discernment of where to place the exact “corner” of a unit became difficult because of lack of a corner edge or a plume of “corner” pixels protruding from a vertex. This is just a sample of the large variety of pixelated corners that result from the scanning process. Though it could be considered trivial, the small errors within the rectifications using these corners could propagate when applied to the vectorization of the artifacts within them.
For example, in addition to the corner shapes, the patterns and directions of the sides of the borders were good indicators of where the true “line” of the edge would lie as illustrated by Figure 5. By noting the general direction of the skew or warp of the edge lines and their frequency of shifting from one direction to another, a general idea of where the corner should be when two lines meet can be more accurately determined if the trends of each line are taken into account.

**Figure 5**

*Figure 5: Seen both in the context of the entire edge as well as a closer look at one of its sections, the edge here clearly steps down at a regular interval of pixels. It is from cues such as these that the determination of the end point of an obscure corner such as those seen in Figure 4 can be objectively determined by what level of the “steppe” it falls onto coming from each edge leading to it.*

This was especially important in cases in which the scanned image contained a large number of pixels or if the borders of the map itself were drawn by hand as seen in Figure 6. When dealing with maps such as these the use of digital gridlines or even a ruler or straight edge up against the computer screen can be used to great effect. By noting the number of pixels of each square in the grid, the outside and inside borders can be determined by counting the number of pixels of each square to the edge of the map. In addition to this, the pixel line of the edge, once found, can be followed down to the opposite side of the map thus resulting in one straight side of the map and the other three can be done in a similar manner to find the true corners of the map at which to place the ground control points. This is also another illustration of how careful alignment during the scanning process can help to avoid warping of edges which can affect the angle the edges follow though admittedly this is always difficult to avoid.
Figure 6: Note how the columns of pixels can be followed all the way down the image. It is more important to be consistent throughout the edge following process to find corners in images such as these than it is to find the “correct” corner to start from, though both are ideal.

One final recurring factor that will help to determine edges in this way is the “baseline”. As seen in Figure 7 below, protrusions of one to several pixels can extend from the primary line which the edge follows. In this study, these were not considered to be part of the edge but rather an indication of a coming shift in its direction. Looking at the frequency of these protrusions, it is easier to get an idea of how soon and at what frequency the direction of the line shifts one pixel over because the protrusions from the previous baseline become the next baseline. It is important to pay attention to the common trend of all the protrusions of an edge and even of the entire map rather than just a few because some edges contain them going in both directions. Essentially a baseline can be defined by the fact that it has protrusions coming from it, and a new baseline is established when a large portion of protrusions begin coming from the old protrusions.
Figure 7: Notice that despite the definitions of baselines and the directions of the edges, this particular corner is still difficult to place. At the end of the y-axis, the protrusions and the baseline are no longer distinguishable which could be interpreted as the baseline being on the outside corner. However, the protrusions indicate that the outside corner of the baseline is actually in the middle of the thick black line. Though subjective, the “correct” edge can be determined based upon the x-axis edge. It is important that the operator use the same decision making style and rules for situations like these throughout the vectorization process.

As a best practice, corner shape, direction of edges, and frequency of edge direction shift should be taken into account when georeferencing irregular scanned maps. Of course this can be avoided with a controlled and consistent scanning practice as well as a higher scanning resolution, but even in the best cases errors are likely to occur and a repeatable standard is required so that they can be dealt with consistently. By following this protocol, all scanned unit images were rectified in the same way and the root means square error of all rectifications was kept at less than one pixel. Though it may seem trivial, having the utmost accuracy in this stage of ground control point placement and georeferencing can noticeably affect later results during
the vectorization process. Errors here will affect errors in the vectorization of artifacts and from there any operations done with the resulting artifacts will continue to propagate the total error.

In most cases, georeferencing software will require the manual placement of three ground control points to calculate a model to automatically determine the location of additional points. As a final verification that the ground control point placement was done properly, the fourth corner point can be deleted from the reference and replaced manually which will cause the model to automatically place a fourth corresponding ground control point on the scanned unit image. If the point is placed at or near the location determined appropriate by the above protocol, the other points are at the correct location. This was done while georeferencing all units in this study to ensure accuracy.

Once all of the rectifications were complete, the unit images could be loaded into the ArcMap (Esri 2015) document in alignment with the shapefile of the excavation grid, an example of which can be seen in Figure 8 though not all rectified images could be shown due to the sketches of the same unit at multiple levels of depth overlapping one another. With all of the hand drawn unit maps in place, the manual vectorization could begin. To regulate this, a set of rules was created. Like the rules made for the determination of the unit corners in the georeferencing process, the approach was intended to be simple, replicable, and conservative so as not to accidentally exclude any part of an artifact’s border.

Figure 8

A new shapefile within a new file geodatabase was created to hold the polygons that would be traced over the artifact drawings. The WPC local coordinate system was used for each
and the x, y tolerance was set to the lowest possible setting so that drawn polygon vertices would be less likely to “snap” to another location if they were drawn lower than the tolerance. This way the highest accuracy could be maintained while drawing the vectors.

To perform the vectorization, the entire frame of view would be zoomed to contain the entire pixelated artifact (or a section of it in the case of large artifacts); the goal being to be able to see the individual pixels of the scanned unit image. Based upon the same principles as the georeferencing process, the trace was to be conservative and assume that to some degree the borders of the actual drawings extended slightly beyond the black pixels. With this in mind, a vertex was placed at each outside corner pixel of the artifact. A series of pixels at the same x or y with a gap in between them or a dip of only 1-2 pixels in size were treated as a straight line and the gap or dip was skipped over as seen in Figure 9.

Figure 9
Illustrated in Figure 10, some polygons included irregular marks extending from their sides. Initially in such cases, the original documents were examined until it became easy to discern these markings as erroneous marks made during the original drawing of the maps which were not intended to be a part of the artifact being drawn. Referencing the original maps made these identifiable in cases in which this was not obvious and in all cases the erroneous marks were passed over while vectorizing.

If an artifact crossed the border of a gridline or fit up alongside it, it was assumed that the artifact ended at the border unless a clear set of pixels not belonging to the regular variations in the gridline borders was present on the other side. Examples of this can be seen at the top of Figure 9 as well as in Figure 10 and Figure 11.

Figure 10
Cases unique to the dataset also arose. For example, one artifact was not drawn at all but instead only an arrow pointing to its position and a label of its number was written. In this case, one pixel at the tip of the arrow was traced since it could only be assumed that the location was exact. If an artifact was drawn sharing a border with another artifact or with a feature of the landscape (such as a boulder), then the inside pixels of the opposing object were used as the outside pixels of the object being traced. For example, when tracing artifact X that shared a border with artifact Y, the inside pixel edges of artifact Y along the shared border were treated as the “outside” pixel edges of artifact X and vice versa. An example of this can be seen in **Figure 12**.
When the pencil markings of a particular unit were soft and the artifact outlines were sparse, the outmost pixels surrounding the artifact were used as its border, skipping the various
gaps in between them to create the largest surface area possible so as not to accidentally cut off any part of the artifact. However, if black pixels in the scan were more than one pixel away from the main body, they were not included as part of the object. In any case in which the outline of an artifact contained smaller objects or lines within it, the outermost layer was always treated as the actual border. An example of this can be seen in Figure 13.

**Figure 13**

When all artifacts were traced and thus transformed into vector polygons, the result was a large sample size spanning a majority of the area of the excavation site as seen in Figure 14. Recall that the artifacts shown are from all levels of all the useable units from all field seasons and so are overlapping and collapsed. As an accuracy assessment of this entire process, a near operation was performed between each artifact polygon and its “true” location indicated by the theodolite points taken in the field. This was done both to the nearest edge of the artifact polygon as well as to the centroid (center-point) of each polygon. This is because the centroids would account for any error in the vectorization of the polygon edges because such irregularity would not greatly affect the centroid’s position. The centroids were generated not being constrained by their polygon’s borders as some of the artifacts were drawn as several distinct pieces. These two datasets would help to get a general idea of the “range” in which the true near distance between the points and polygons would lie.
Because the polygons and points were so closely grouped together in each unit (due to each grid containing the artifacts of multiple seasons that often overlapped in x, y location), running a blind near operation between the two files resulted in many errors and mistakes. To remedy this and save time, a more sophisticated model was made to ensure that each matching point and polygon or point and centroid pair would be identified before an iterative near operation would take place. To do this, a unique identifier was given to each point, polygon, and centroid consisting of their unit location and artifact number combined. This way only one of each group shared the same unique ID and so could be identified and the near distance between them calculated for them only. For example, artifact 49 in unit M98 would have a unique id of M98_49 and the point, polygon, and centroid representing that artifact would be the only features identified with that unique ID.

Based upon this, an iterative near operation for each unique pair was made into a model in ArcMap (Esri 2015) with one model performing the operation between the “true” points and the polygons and a second performing the operation between the “true” points and the centroids of each polygon. Figure 15 shows how the “true” point and polygon/centroid of a unique ID would be selected and made into two individual outputs upon which the near operation would be performed. The iterator would then select the next unique ID and the process would be repeated for all possible unique IDs in the list. This approach is not only much faster than manually selecting out the features containing each unique ID but it also allows no room for a mistaken near operation since the only features the near distance will be calculated for must share the unique ID value.
With the artifacts converted to vectors and the near operations complete, the output files of each individual unique ID point containing the near distances could be joined and the resulting data exported to a table from which statistics could be drawn. These statistics then give an idea not only of the accuracy of the vectorization process itself but also an interesting look at the accuracy of the excavator’s drawings on the map.

The full process from start to finish can be explained simply in a few key steps:

1. Copying artifact map images from pdf or other file types, or scanning original maps; either of which must be done at the highest resolution possible.

2. Alignment of these individual maps with their respective units by either tying the maps together in series or referencing a vectorized template of their layout and using ground control points to mark common locations on each for rectification. During this process, careful attention must be paid to the baseline, protrusion, and corner patterns of each individual map.

3. Conservative tracing of the outside edges of the outside pixels of each artifact drawn to the map and saving each trace as a part of a new shapefile or other vector format.

4. If theodolite-derived points or other reference exist for the traced artifacts, an iterative near operation can be performed to get an idea of how accurate the tracings are to the “true” locations represented by the reference (however this is not telling of how well the actual tracing was done).
Results

Using the vectorized polygons, their centroids, and the “true” theodolite-derived points, many insightful operations can be done and an accuracy assessment can be made of the manually vectorized artifacts. In the first part of the results (up to and including Figures 16-20), data is given that can be used as an accuracy assessment comparing the “true” theodolite-derived points with the polygon centroids (center-points) and nearest polygon edge to them, respectively. In the second part (Figures 21-26) examples are given of operations that can be performed with the new data created from using this methodology. The results of each are evaluated in more detail in the discussion section.

Accuracy Assessment

The statistics derived from the polygons alone is enough to offer an idea of how the different groups of data relate to one another and could be useful in identifying any problem areas or sources of error. For example, in both Figures 16 and 17, FCR (Fire Cracked Rock), Lithic, Shell, and Bone (in that order) hold the highest average error for groups with a count greater than 10. Interestingly, these four groups also have the widest ranges while ceramic has a very low average error and narrow range in spite of its high count.

Figures 18-20 go a step further than the tables and actually show the distribution of the entire datasets in relation to their standard deviations. Note how the data of both groups is similar and somewhat resembles a normal distribution but leans below and at the mean. In the context of this dataset, this means overall a low rate of error in the field drawings despite several stark outliers in the upper end. However, the mean still sits just above 10 centimeters which could or could not be a tolerable level of error depending upon the goals of the project the drawings would be used for. Though the tolerable level of error is the ultimate decision of the project leader, this study arbitrarily considers 5cm (-0.375 SD’s from mean for polygons, -0.493 SD’s from mean for centroids) as the tolerable error threshold with 48.52% of polygons and 34.91% of centroids falling at or below this.
<table>
<thead>
<tr>
<th>Polygons</th>
<th>SD (As Sample)</th>
<th>Count</th>
<th>Group</th>
<th>Average</th>
<th>Min</th>
<th>Max</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1681</td>
<td>338</td>
<td>All</td>
<td>0.1130</td>
<td>0</td>
<td>1.0373</td>
<td>0.0557</td>
</tr>
<tr>
<td></td>
<td>0.0849</td>
<td>4</td>
<td>Charcoal</td>
<td>0.1377</td>
<td>0.0185</td>
<td>0.2182</td>
<td>0.1570</td>
</tr>
<tr>
<td></td>
<td>0.0946</td>
<td>82</td>
<td>Ceramic</td>
<td>0.0781</td>
<td>0</td>
<td>0.0059</td>
<td>0.481</td>
</tr>
<tr>
<td></td>
<td>0.1506</td>
<td>94</td>
<td>Bone</td>
<td>0.1039</td>
<td>0</td>
<td>0.9661</td>
<td>0.0485</td>
</tr>
<tr>
<td></td>
<td>0.2223</td>
<td>29</td>
<td>FCR</td>
<td>0.2260</td>
<td>0</td>
<td>0.8579</td>
<td>0.1236</td>
</tr>
<tr>
<td></td>
<td>n/a</td>
<td>2</td>
<td>Glass</td>
<td>0.1145</td>
<td>0.0635</td>
<td>0.1655</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>0.2017</td>
<td>56</td>
<td>Lithic</td>
<td>0.1132</td>
<td>0</td>
<td>1.0336</td>
<td>0.0403</td>
</tr>
<tr>
<td></td>
<td>n/a</td>
<td>1</td>
<td>Microfauna</td>
<td>0.2707</td>
<td>n/a</td>
<td>n/a</td>
<td>0.0408</td>
</tr>
<tr>
<td></td>
<td>0.0337</td>
<td>61</td>
<td>Ochre</td>
<td>0.0408</td>
<td>0.0095</td>
<td>0.0957</td>
<td>0.0957</td>
</tr>
<tr>
<td></td>
<td>0.0408</td>
<td>1</td>
<td>Historic</td>
<td>0.0271</td>
<td>n/a</td>
<td>n/a</td>
<td>0.0095</td>
</tr>
</tbody>
</table>

**Figure 16:** Notice how some of the lowest errors include ceramic while the highest error again only comes from FCR, Lithic, Shell, and Bone.
### Centroids

<table>
<thead>
<tr>
<th>Group</th>
<th>SD (As sample)</th>
<th>Count</th>
<th>Count</th>
<th>Group</th>
<th>SD (As sample)</th>
<th>Count</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.1695</td>
<td>338</td>
<td>4</td>
<td>Charcoal</td>
<td>0.0802</td>
<td>4</td>
<td>82</td>
</tr>
<tr>
<td>Ceramic</td>
<td>0.0969</td>
<td></td>
<td></td>
<td>Ceramic</td>
<td>0.0979</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Bone</td>
<td>0.1511</td>
<td>94</td>
<td>29</td>
<td>FCR</td>
<td>0.2224</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCR</td>
<td></td>
<td></td>
<td></td>
<td>Glass</td>
<td>0.1880</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glass</td>
<td></td>
<td></td>
<td></td>
<td>Ochre</td>
<td>0.0346</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>Ochre</td>
<td></td>
<td></td>
<td></td>
<td>Historic</td>
<td>0.1290</td>
<td>1</td>
<td>61</td>
</tr>
<tr>
<td>Historic</td>
<td>n/a</td>
<td>1</td>
<td>61</td>
<td>Unknown</td>
<td>0.10476</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>n/a</td>
<td>1</td>
<td>61</td>
<td>Unknown</td>
<td>n/a</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 17**
Figure 18

**Centroid Error Frequency**

- Frequency values for standard deviations from mean:
  - -0.75 to -0.625: 65
  - -0.625 to -0.5: 40
  - -0.5 to 0: 43
  - 0 to 0.5: 43
  - 0.5 to 1: 28
  - 1 to 2: 7
  - 2 to 3: 1
  - 3 to 4: 3
  - 4 to 5: 2

**Polygon Error Frequency**

- Frequency values for standard deviations from mean:
  - -0.625 to -0.5: 70
  - -0.5 to 0: 48
  - 0 to 0.5: 37
  - 0.5 to 1: 51
  - 1 to 2: 26
  - 2 to 3: 6
  - 3 to 4: 6
  - 4 to 5: 1
  - 5 to 6: 3
Figure 19

**Centroid Standard Deviations from Mean**

**Polygon Standard Deviations from Mean**
Figure 20

Polygon Standard Deviation from Mean

Centroid Standard Deviation from Mean
Examples of Use

**Figure 21** gives a unique look at potential sources of error by identifying the artifacts with the highest error in the entire dataset in comparison with those with the lowest amount of error. These juxtapositions help to identify problem units or problem artifacts which might deserve a second look and also further support the patterns shown in the tables. Despite these seemingly extreme outliers, the entire dataset was actually continuous in its pattern with error levels changing gradually throughout.

These patterns are also present in the full list of artifact drawings (and thus their vector polygons) which contained no error as shown in **Figure 22** below. Ceramic makes a frequent appearance on the list as do FCR, Bone, Shell, and Lithic (Groundstone included) which further shows the extent of their varying range of error. In addition to this the zero list helps to identify units in seasons which contained these correct artifacts so that any investigator would have reference points to look into why certain units may contain less error in their artifact records and sketches than others.

Information such as the total area of artifacts found can be valuable and by creating three dimensional renderings of artifacts *in situ*, the total area of overlapping artifacts can be determined along with the area of visible concentrations of material. The average size of artifacts in an entire site can also be calculated along with other summary statistics like those seen in **Figure 23** which can be revealing in regards to a site’s history and evolution (a larger average artifact size could indicate a lack of disturbance for instance). Though these estimations are in two dimensions rather than three, these shapefiles can provide a platform for later creation of three dimensional shapefiles depending upon available data.

**Figure 21**

<table>
<thead>
<tr>
<th>Polygons</th>
<th>Centroids</th>
</tr>
</thead>
<tbody>
<tr>
<td>10Highest</td>
<td>10Lowest(NoZero)</td>
</tr>
<tr>
<td>1.037269 BONE</td>
<td>0.685474 FCR</td>
</tr>
<tr>
<td>1.033607 FCR</td>
<td>0.693262 FCR</td>
</tr>
<tr>
<td>0.96614 LITHIC</td>
<td>0.738336 LITHIC</td>
</tr>
<tr>
<td>0.857935 SHELL</td>
<td>0.767287 SHELL</td>
</tr>
<tr>
<td>0.780663 SHELL</td>
<td>0.790629 SHELL</td>
</tr>
<tr>
<td>0.758192 LITHIC</td>
<td>0.791548 LITHIC</td>
</tr>
<tr>
<td>0.738406 FCR</td>
<td>0.872243 FCR</td>
</tr>
<tr>
<td>0.715502 BONE</td>
<td>0.986474 BONE</td>
</tr>
<tr>
<td>0.674012 LITHIC</td>
<td>1.047556 Unknown</td>
</tr>
<tr>
<td>0.64274 Unknown</td>
<td>1.056662 LITHIC</td>
</tr>
</tbody>
</table>

*Figure 21: The values shown here are the near distances in meters. Some of the lowest errors in either dataset are one millimeter or less while some of the highest can be as much as one meter.*
<table>
<thead>
<tr>
<th>YEAR</th>
<th>UNIT</th>
<th>ARTIFACTTYPE</th>
<th>Stratum</th>
<th>Art_ID</th>
<th>UnID</th>
<th>NEAR_DIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>L96</td>
<td>BONE</td>
<td>1</td>
<td>9</td>
<td>L96_9</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>L96</td>
<td>FCR</td>
<td>1</td>
<td>18</td>
<td>L96_18</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>L96</td>
<td>SHELL</td>
<td>1</td>
<td>19</td>
<td>L96_19</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>L97</td>
<td>BONE</td>
<td>3</td>
<td>18</td>
<td>L97_18</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>L97</td>
<td>CERAMIC</td>
<td>1</td>
<td>3</td>
<td>L97_3</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>L97</td>
<td>LITHIC</td>
<td>1</td>
<td>4</td>
<td>L97_4</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>L98</td>
<td>SHELL</td>
<td>2</td>
<td>2</td>
<td>L98_2</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>M97</td>
<td>BONE</td>
<td>1</td>
<td>29</td>
<td>M97_29</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>M98</td>
<td>BONE</td>
<td>1</td>
<td>2</td>
<td>M98_2</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>K95</td>
<td>LITHIC</td>
<td>3</td>
<td>24</td>
<td>K95_24</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>K96</td>
<td>BONE</td>
<td>3</td>
<td>18</td>
<td>K96_18</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>K96</td>
<td>BONE</td>
<td>3</td>
<td>18</td>
<td>K96_18</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>K96</td>
<td>GROUNDSTONE</td>
<td>3</td>
<td>20</td>
<td>K96_20</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>K97</td>
<td>SHELL</td>
<td>3</td>
<td>10</td>
<td>K97_10</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>K97</td>
<td>SHELL</td>
<td>3</td>
<td>11</td>
<td>K97_11</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>K97</td>
<td>SHELL</td>
<td>3</td>
<td>9</td>
<td>K97_9</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>L95</td>
<td>CERAMIC</td>
<td>3</td>
<td>33</td>
<td>L95_33</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>M98</td>
<td>BONE</td>
<td>3</td>
<td>17</td>
<td>M98_17</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>J97</td>
<td>CERAMIC</td>
<td>5</td>
<td>11</td>
<td>J97_11</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>K97</td>
<td>FCR</td>
<td>5</td>
<td>31</td>
<td>K97_31</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>K97</td>
<td>LITHIC</td>
<td>5</td>
<td>30</td>
<td>K97_30</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>L97</td>
<td>CERAMIC</td>
<td>3</td>
<td>22</td>
<td>L97_22</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>M97</td>
<td>BONE</td>
<td>3</td>
<td>52</td>
<td>M97_52</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>M97</td>
<td>CERAMIC</td>
<td>5</td>
<td>61</td>
<td>M97_61</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>M97</td>
<td>LITHIC</td>
<td>5</td>
<td>62</td>
<td>M97_62</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>H98</td>
<td>BONE</td>
<td>5</td>
<td>17</td>
<td>H98_17</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>H98</td>
<td>BONE</td>
<td>3</td>
<td>4</td>
<td>H98_4</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>I96</td>
<td>LITHIC</td>
<td>5</td>
<td>14</td>
<td>I96_14</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>I97</td>
<td>CERAMIC</td>
<td>3</td>
<td>5</td>
<td>I97_5</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>I97</td>
<td>SHELL</td>
<td>3</td>
<td>4</td>
<td>I97_4</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>J96</td>
<td>LITHIC</td>
<td>5</td>
<td>19</td>
<td>J96_19</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>K95</td>
<td>CERAMIC</td>
<td>5</td>
<td>35</td>
<td>K95_35</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>K95</td>
<td>CERAMIC</td>
<td>5</td>
<td>36</td>
<td>K95_36</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>K95</td>
<td>CERAMIC</td>
<td>5</td>
<td>38</td>
<td>K95_38</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>K95</td>
<td>SHELL</td>
<td>5</td>
<td>39</td>
<td>K95_39</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>K95</td>
<td>SHELL</td>
<td>5</td>
<td>40</td>
<td>K95_40</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>K95</td>
<td>SHELL</td>
<td>5</td>
<td>42</td>
<td>K95_42</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>L96</td>
<td>CERAMIC</td>
<td>5</td>
<td>98</td>
<td>L96_98</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>L96</td>
<td>CERAMIC</td>
<td>5</td>
<td>99</td>
<td>L96_99</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>L96</td>
<td>LITHIC</td>
<td>5</td>
<td>101</td>
<td>L96_101</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>L96</td>
<td>LITHIC</td>
<td>5</td>
<td>94</td>
<td>L96_94</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>L96</td>
<td>SHELL</td>
<td>5</td>
<td>96</td>
<td>L96_96</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>L96</td>
<td>SHELL</td>
<td>5</td>
<td>97</td>
<td>L96_97</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>M95</td>
<td>LITHIC</td>
<td>5</td>
<td>39</td>
<td>M95_39</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 23

<table>
<thead>
<tr>
<th>Areas of Artifact Polygons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>0.0013</td>
</tr>
</tbody>
</table>

Though the statistical analyses done on such a dataset can be powerful, they are even more useful when displayed and investigated using GIS software. Figure 24 shows the difference between the original illustrations used to depict the artifact spatial organization of each unit and the newly vectorized artifact polygons which act as a supplement to the original records rather than a replacement. The level of detail available for analysis increases significantly with the original illustrations offering limited spatial analyses while the dataset in a GIS allows for a multitude of complex selections and geometric or distances measures using both generated data tables and the maps themselves.

Figures 25 and 26 show an example of how this can be explored in greater depth. For instance, by adding the attribute of percentage of artifacts within a certain range of error to an abstraction of the excavation area, the patterns in the dataset can be seen as the error tolerance is changed and multiple seasons compared. Though the data itself cannot identify why certain patterns appear, it can help identify them for further investigation.

Figure 24

Original Illustration

Visual Analysis

(Enloe 2012-2015b)

(Enloe 2015)
Figure 24: While the visual analyses possible with illustrations containing little to no spatial information can be telling of spatial patterns, more complex algorithmic searches and comparisons can be done with data such as vectorized polygons which contain spatial data in a coordinate system as shown on bottom.
Figure 25: Note that the percentage here was determined by the total number of artifacts found within each particular unit. Notice the higher percentage rate near the units in the upper left.
Figure 26: Unlike Figure 25, the percentage here was determined by the total number of artifacts that could be traced from the entire season and the darker units represent those which contained the most accurate artifact drawings (those that are closest to their “true” value as referenced by the theodolite points.)
Discussion

Though the uses of data here are not exhaustive they are very telling of the many benefits of the vectorization of artifact data into polygons. Firstly, new information such as total area of artifacts and excavation related errors or problem areas can be derived. On top of this, vectorized artifact data can serve as a secondary and complimentary dataset to both Theodolite point data and the original illustrations created for a site. For example, when theodolite points are taken, the location on the artifact where the prism (used to capture and reflect the laser used by the theodolite) is placed is not often recorded which could result in small errors that could accumulate and become worse if such data is used in a GIS. The polygon shapefiles act as a secondary reference in relation to these to give an idea of the prism placement as well as give an idea of the overall shape of the artifact.

The polygons can be used to test previous interpretations of a site artifact spatial array by adding additional samples to the analyzed dataset and even act as a platform for new and more appealing visualizations such as three dimensional shapefiles for use in publications and presentations. Generally in the context of archaeology more data and reference points are always useful because the actual site itself cannot be un-excavated or brought back. However, these are just a few of the possible ways in which data from previously excavated sites can itself be excavated and retrieved for the benefit of the project.

The methods provided here are a simple, easy to follow way to access these potential benefits and while every site and case is slightly different, the guidelines described here can be referenced for any project that requires low error vectorization of a scanned image, particularly if it is done by the same individual in a manner that fine-tunes these methods to the particularities of a specific project. Creating and following a decision tree or rules that elaborate on those presented here that are specific to the needs of each individual case can result in extremely acceptable results.

It is still important to consider that these methods are manual which will always to some degree imply subjectivity, human error, and a high cost in time and energy. The reason for a need for standards for manual vectorization are due to the fact that there is currently no effective automatic method to use for irregular maps and datasets. This is especially true for historical and hand drawn maps such as those used here. Many teams realize that manual vectorization is still the most efficient and accurate way to prepare historical maps for GIS analysis but none mention the specific methods for or the reasoning behind the way in which it was done. This then leads to a potential for irregular practice and thus irregular results while a common practice would lead to more mutually understandable results without the need to elaborate.

Of course the manual method is even considered by those that use it to be a less than ideal way to achieve a vectorized end product. The types of historic and hand-drawn maps which would require more specific algorithms and automatic vectorization programs are not often given much attention due to the general rarity of such projects in relation to the overwhelming need for algorithms for industrial, topographic, and road maps. Those that are interested in a better way to
vectorize complex historic maps do not have the skillsets necessary to develop it while those that do are often not concerned, uninterested, or unaware of this need.

There are many directions in which this growth could take place. Balaji and Punithavalli (2010) offer a supervised and more object-detection based method that would be more suitable to historic datasets. Utilizing learning algorithms and training data taken from historic examples would help build upon already existing techniques that could be fashioned to be used with these sorts of maps. In addition Yang (2012) mentions a tracing algorithm that follows pixel corners much in the same way the manual method presented here does albeit for unbroken lines only. If such an algorithm could be improved or changed to be used specifically with discontinuous features, it would be much more useful in the historical context. These are but a few of the possibilities for future development and with the proper manipulation of current methods, an automatic technique may soon be able to replicate the methods shown in this study.

Conclusion

Though the automation of digitizing and vectorization of complex historic and hand-drawn maps has not yet been realized, many projects are underway which require it. This has required the use of manual methods for which there were no universally recognized or available standards and this study has made an attempt at establishing guidelines for this process which are universal, repeatable, and simple. The results of this process can be used as a geospatial supplement to already existing site data which provides a more complex geospatial element to the site analysis and enhances estimations and interpretations gleaned from that dataset.

Despite this benefit, a more automated method is desired to increase efficiency and help to preserve historical documents more accurately and at a faster rate to be used in many areas of study. By defining a set of rules to be followed until such a method is developed, this study also provides a basic workflow for what these automatic methods need to accomplish. This way even more information and previously unknown data can be gleaned from the historic resources at hand by putting them in a form that can be utilized by the technological prowess of the modern era.

References

Adobe Systems Incorporated
2016 Adobe Photoshop CC. Computer Software. San Jose, California.

Balaji, N.V., and M. Punithavalli

Caldwell, Warren, W.
Chazan, Michael

Chiang, Yao-Yi
2010 Harvesting Geographic Features From Heterogeneous Raster Maps. Ph.D. dissertation, Department of Computer Science, University of Southern California, Los Angeles.

Daniels, Richard C., and Robert H. Huxford

Daryal, Neeti

Dharmaraj, Girija

Enloe, James G.
2012-2015a Woodpecker Cave Field Dossiers. Manuscripts on file, Department of Anthropology, University of Iowa, Iowa City, Iowa.

2012-2015b Woodpecker Cave Report No. 1 – 4. Manuscripts on file, Department of Anthropology, University of Iowa, Iowa City, Iowa.


Esri (Environmental Systems Research Institute Inc.).


Hexagon A.B.  

Kaim, Dominik, Jacek Kozak, Natalia Kolecka, Elzbieta Ziołkowska, Krzysztof Ostafin, Katarzyna Ostańowicz, Urs Gimmi, Catalina Munteanu, and Volker C. Radeloff  

Mabee, S.B., R. Newton, A. Stepanov, D. Ene, and D. Ivanov  

McGrath, James  
2015a *WPC_2015_Grid*. Data on file, Department of Anthropology, University of Iowa, Iowa City, Iowa.  
2015b *wpcPlotsall*. Data on file, Department of Anthropology, University of Iowa, Iowa City, Iowa.

Nobajas, Alexandre, and Francesc Nadal  
2015 From historical map to online 3D recreation: the 1861 cadastral map of Horta (Barcelona). *Cartography and Geographic Information Science* 42:211-223.

Oka, Shriram, Akash Garg, and Koshy Varghese  

Rapson, David  
1990 *Pattern and process in intrasite spatial analysis: Site structural and faunal research at the Bugas-Holding site*. Ph.D. Dissertation, The University of New Mexico, Albuquerque.

Schiffer, Michael  

Schneider, G. Michael, and Judith L. Gersting  

Schuppert, Christof, and Andreas Dix  
Sharma, Ojaswa

Skalos, Jan, Martin Weber, Zdenek Lipský, Ivana Trpáková, Markéta Santrucková, Lenka Uhlírová, and Pavel Kukla

Walls, Michael. D.

Wiseman, James, and Farouk El-Baz

Wu, Chi T.
1998 *A study on Positional Accuracy of Scanned and digitized topographic maps*. Ph.D. dissertation, Department of Civil Engineering, National Chung Hsing University, Taichung, Taiwan.

Yang, Y., An, X., Huang, L.