Investigative Analysis across Documents and Drawings: Visual Analytics for Archaeologists

V. Deufemia, L. Paolino, G. Tortora  
Università di Salerno  
Via Ponte don Melillo  
84084 Fisciano(SA), ITALY  
+39 089963346  
{deufemia,ipaolino,tortora}@unisa.it

A. Traverso  
Lab. di Paletnologia, Università di Genova  
Via Balbi 4  
16126 Genova, ITALY  
+39 0102099712  
antonella.traverso@lettere.unige.it

V. Mascardi, M. Ancona, M. Martelli  
Università di Genova  
Via Dodecaneso 35  
16146 Genova, ITALY  
+39 0103536664  
{mascardi,ancona,martelli}@disi.unige.it

N. Bianchi, H. De Lume  
Lab. dép. de préhistoire du Lazaret  
33 Bis Boulevard Franck Pilatte  
06300 Nice, France  
+ 33 0493568179  
{nbianchi,lazaret}@lazaret.unice.fr

ABSTRACT

With the invention and rapid improvement of data-capturing devices, such as satellite imagery and digital cameras, the information that archaeologists must manage in their everyday’s activities has rapidly grown in complexity and amount. In this work we present Indiana Finder, an interactive visualization system that supports archaeologists in the examination of large repositories of documents and drawings. In particular, the system provides visual analytic support for investigative analysis such as the interpretation of new archaeological findings, the detection of interpretation anomalies, and the discovery of new insights. We illustrate the potential of Indiana Finder in the context of the digital protection and conservation of rock art natural and cultural heritage sites. In this domain, Indiana Finder provides an integrated environment that archaeologists can exploit to investigate, discover, and learn from textual documents, pictures, and drawings related to rock carvings. This goal is accomplished through novel visualization methods including visual similarity ring charts that may help archaeologists in the hard task of dating a symbol in a rock engraving based on its shape and on the surrounding symbols.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – graphical user interfaces (GUI). H.2.8 [Database Management]: Database Applications – data mining. H.5 [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

General Terms

Design, Human Factors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AVI’12. May 21–25, 2012, Capri Island, Italy
Copyright © 2012 ACM 978-1-4503-1287-5/12/05.. $10.00

Keywords

Visual analytics; information visualization; investigative analysis; rock art archaeology.

1. INTRODUCTION

In the recent years, the amount and variety of data accessible to archaeologists has grown exponentially. This is motivated, on the one hand, by the rapid growth of devices for delivering vast amounts of data, such as satellite and geophysics imagery, laser scanners, GPS, handheld XRF; on the other hand, by the possibility of accessing data-rich sources of archaeological information. Examples of initiatives in this direction are the development of open access digital archives such as the Archaeology Data Service [2], the Open Context [30], and the Digital Antiquity [28][33], and the development of widely accepted ontologies to describe archaeological information [31].

With the increasing complexity of massive archaeological data, the need to support archaeologists in the inherently analytical nature of their investigations increases as well. Indeed, although many methods have been developed and improved for collection, analysis, visualization, and modeling of archaeological site information, included the use of virtual reality [3], little efforts have been devoted to support the analytical activities of archaeologists in their everyday’s work and when they operate on the field. Thus, the high-dimensionality of data and the type of investigations archaeologists perform, suggested us to develop new forms of visualization for providing an integrated environment where archaeologists can investigate, discover, and learn from both textual documents and pictures.

In this paper we present Indiana Finder, an interactive visualization system that supports archaeologists in the examination of large repositories of documents and drawings. In particular, the system provides visual analytic support for investigative analysis such as the interpretation of new archaeological findings, the detection of interpretation anomalies, and the discovery of new insights. This is accomplished through novel visualization methods including visual similarity ring charts that may support archaeologists in dating a symbol in a rock carving based on its shape and on the surrounding symbols.
This research is part of the Indiana MAS project\(^\dagger\) whose main purpose is to provide a framework for the digital protection and conservation of rock art natural and cultural heritage sites, by storing, organizing and presenting information about them in a systematic way. In this domain, Indiana Finder allows archaeologists to investigate textual documents, pictures, and drawings related to rock carvings.

Indiana Finder, the first component of the Indiana MAS system, has been prototyped and tested on data from the Adevrepam database containing information about all the rock carving reliefs of Mount Bego (45,000 records). Each record of the database consists of textual interpretation description, ontological information, geographical coordinates, pictures and drawings of the reliefs. The prototype supports archaeologists in their investigation activities by performing analytics on the dataset, considering the geographical context and the relationships with other findings recovered in the area, and by providing appropriate visualization and interaction techniques.

The remainder of this paper is organized as follows. In Section 2, we present background information on digital protection and conservation of rock art natural and cultural heritage sites. Section 3 presents related work in visual analytics and visualization of archaeological information. Sections 4 and 5 illustrate the proposed system and three application scenarios. Section 6 discusses conclusions and presents ideas and plans for future work.

2. ROCK ART ARCHAEOLOGY

Our work is situated in the domain of rock art archaeology where we worked with prehistorian archaeologists involved in the study of rock carving sites.

Rock art research started to gain more and more attention at the turn of the millennium, when many scientific publications on the topic appeared [4][11][12][20][39]. A federation of national and regional organizations promoting the study of palaeoart and cognitive archaeology, IFRAO\(^\dagger\), have been existing for more than 20 years.

Despite all this effort and the successful results achieved in the last years, archaeologists would still appreciate a support in analyzing the implications of rock art in a systematic and objective way. Indiana Finder, aiming at providing archaeologists with such a support, will be evaluated on data coming from Mount Bego (southeastern France), one of the richest and most important European rock art sites.

Mount Bego was described as “a hellish place with pictures of demons and a thousand devils cut everywhere on the rocks” by Pierre de Montfort, a French voyager who visited that area in the XV century. The superstition around this place must have been deep-rooted, as the toponymy of the surrounding areas still has evident references to a hellish world: Cime du Diable (Peak of the Devil), Lacs de l’Enfer (Lakes of the Hell) and Vallée de la Sorcière (Valley of the Witch). Far from being a hellish place, Mount Bego is an invaluable source of information on everyday’s life of ancient populations. Archaeologists have counted 35,814 signs made by pecking in the area [5]. Carved rocks are situated in the high valleys of what is now the Mercantour National Park, over 2000 meters of altitude. No scientific study was carried out on the carved rocks until the end of XIX century, when Clarence Bicknell, British archaeologist and botanist, came to the area with his friend Luigi Pollini and sketched several drawings on small sheets of paper. Between 1898 and 1910 Bicknell realized up to 13,000 drawings and reliefs, part of which were then published in [6]. Bicknell identified seven types of figures taking a natural history approach: horned figures (mainly oxens), ploughs, weapons and tools, men, huts and properties, skins and geometrical forms [10]. Fig. 1 shows a picture of the “wizard” rock carving together with Bicknell’s original relief and drawing, and the digitalized relief made by De Lumley’s equipe.

![Figure 1. “Wizard” engraving: picture (a), picture of the relief made by Bicknell on botanic sheet (b), drawing made by Bicknell on his notebook (c), digitalized relief made by De Lumley’s equipe (d).](image)

The division of Mount Bego’s area into sectors, zones, and groups was established between 1927 and 1942 by Carlo Conti, who was in charge of a systematic fixing and copying of the engravings of the region by Piero Barocelli, Superintendent of Antiquities for Liguria, Piemonte and Val d’Aosta. Henry de Lumley followed Conti’s repartition and rock numbering, when, in 1967, he revitalized the research on the site. Today, de Lumley’s team is still active in the area, and bases its work on the same spatial organization. According to De Lumley and Echassoux [14], Mount Bego’s engravings are ideograms bringing out a conception of the world which is transmitted from generation to generation by using a graphic code. It is unquestionable, in fact, that the engravings are strongly related to one another, which is proved by the consistency of the subjects as well as the techniques used to carve them. Because of the strong relationships among Mount Bego’s carvings they represent a good testbed for demonstrating the feasibility of our approach.

3. RELATED WORK

The field of visual analytics (VA) is continuously growing with research efforts expanding into many different domains, such as

---

\(^\dagger\)http://indianamas.disi.unige.it

\(^\dagger\)http://www.cesmap.it/ifrao/ifrao.html
automotive engineering [32] and finance [35]. Here we review related work of existing VA tools supporting investigative analysis and on usage of data visualization and VA in archaeological data analysis.

3.1 Visual Analytics to Support Investigative Analysis
Recent years have seen a rise in the number of research and commercial VA systems built to assist investigative analysis, also called Exploratory Search [26][38]. Information Seeking Support [27], or Sense-making [16][23].

IN-SPIRE is a system for exploring textual data in a collection of documents [42]. The system employs clustering analysis and generates data views based on the mountain and galaxy metaphors to allow analysts identifying trends over time and hot topics. WireVis is a VA tool for identifying specific keywords within financial wire transactions to help detect suspicious activities, as well as possible money laundering among hundreds of thousands of wire transfers [7]. The system employs multiple views on the data including heatmap view, time-series view, a search by example view, and a keyword relation view. ThemeRiver depicts the thematic variations over time in large document collections by using a river metaphor [19]. This metaphor represents themes as colored flowing within the river, which narrow or widen to indicate decreases or increases in the strength of a topic at a specific point in time. TRIST allows analysts to interact with large document collections and to quickly uncover the relevant, novel, and unexpected events [22]. Its user interface provides different perspectives on search results including vector-based clustering, trend analysis, comparisons, and difference. COPLINK is a tool aimed at providing an information extraction facility that integrates information from police case reports and analyzes criminal networks [9]. The system uses data mining techniques to set up a concept space of entities and objects and visualizes crime relationships with a hyperbolic tree view. Jigsaw is a tool aimed at supporting people in the exploration of document collections [34]. The system visualizes connections between entities across documents and provides advanced functionalities such as analysis of document similarity, document sentiment, document clusters by theme or content, and document summarization through a few words or sentences.

To the best of our knowledge none of these tools focus on the investigation of heterogeneous data, such as ontologies and drawings; they instead use structured and unstructured document collections and are more general-purpose VA tools.

3.2 Visualization in the Archaeological Domain
Although data visualization plays a central role in the processes of discovery, analysis, interpretation, and communication, in the archaeology domain it has been mainly employed for illustration and digital archiving [25], or in the context of virtual reality [3].

In the archaeological domain, GIS tools have been equipped with increasing capabilities and employed for decision-support applications [37], for analytical and modeling applications [13][24]. The GIS spatial analytical application presented in [40] allows domain experts to investigate the potential extent of a habitat/environment through the analysis of a series of maps, graphs, and tabular data.

The use of Virtual Reality (VR) in cultural heritage has rapidly grown in popularity in the last few years [3]. The primary focus of the majority of existing VR approaches to cultural heritage is on rendering “realistic” scenes of ancient places that are partially or completely destroyed and to generate holistic experiences of the past [1][18]. Although VR plays a marginal role in the process of archaeological discovery, analysis, or interpretation, it can provide archaeologists new insights by rendering alternative versions of a scene [15], or displaying the results of complex analytical tools [41].

To date, only the work proposed in [21] employs VA in the archaeological context. The authors propose the use of space-time cube, which is a GIS-based implementation of Hägerstrand’s original Spacetime Aquarium [17], for the visualization and analysis of space-time archaeological data. They also present the preliminary results on the development of functions for archaeological investigation within a geovisual analytics environment. Our approach differs from this system in its focus on visually representing relationships between interpretations and pictures in repositories of archaeological findings. We have developed several data visualization interfaces that succinctly represent this information and allow archaeologists to rapidly navigate and access them.

4. INDIANA FINDER
Indiana Finder provides archaeologists with multiple perspectives on an archive of geo-referenced pictures and relics with textual interpretations associated. As shown in Fig. 2, its architecture is composed of three major software components: (1) a Data Preprocessing Component for automatic data filtering, (2) a Data Analysis Component for performing statistical analysis on data, and (3) a Visualization Component for showing summarized results and for gaining insight into correlations among them. The statistical analysis performed on the heterogeneous data includes clustering, correlation, engraving shape analysis, and allows domain experts to detect connections between the entities in the repository. The Visualization component was meant for sense-making activities [23][36] by presenting information about rock carving interpretations and drawings through different views, which are discussed in more detail in the next section.

The Repository is a PostgreSQL database, equipped with the PostGIS module to manage geographical objects and accessed through the PyGreSQL module. The ontology used to organize data associated with rock carvings, included semantic annotations and semantic relationships between them, has been developed on the base of the taxonomy by de Lumley and Echassoux [14]. Fig. 3 shows some meaningful concepts and their generalizations extracted by the Role Ontology Extractor [1]. This ontology will be extended in a semi-automatic way in order to correctly classify multilingual documents and pictures, and not only drawings, and will be used to classify newly inserted documents according to it.

The development of the Indiana MAS system is the long term goal of a recently funded project. This component aims at providing a working multi-agent system supporting archaeologists in many difficult tasks, such as the classification of the relics and their organization within a digital library, the interpretation of data stored in the digital library, the identification of relationships among them, and the enrichment of raw data with semantic information. In this phase of the project, Indiana Finder extracts
data directly from Repository, whereas in the future it will interact with Indiana MAS for collecting semantically richer information, and for providing advanced views on them.

The whole system is written in Java and adopts a model–view–controller architecture that separates data from the user interface components.

5. APPLICATION SCENARIOS

In the following sections, we describe three scenarios where Indiana Finder provides a concrete support to archaeological investigations.

5.1 Finding Places based on Rock Carving Interpretations

The high density of rock carvings in Mount Bego’s area requires advanced interfaces to visually summarize their interpretations and to quickly relate interpretations with sites. As an example, the archaeologist might be interested in reasoning on the concepts that appear most frequently in a particular area of the archaeological site.

The analysis of hundreds textual descriptions representing the interpretations of rock carvings in that particular area of interest is unfeasible. Indiana Finder can represent rock carving interpretations, as well as information on their location, in a compact and intuitive way, which provides nevertheless enough detail to derive comprehensive insights. Such a representation exploits ontology-driven metadata associated with the interpretation of rock carvings, and summarizes them on the archaeological site map.

The analysis process consists of three steps: the automatic extraction of metadata from the ontology, the construction of clusters of metadata, and the construction of layers to be visualized on the map.

**Step 1:** As a first step, the user indicates the area of interest and the level of detail s/he intends to investigate. As an example, if s/he is interested to understand where, within the given area, there are more carvings representing priests (deities, resp.), then s/he has to indicate fine (coarse, resp.) grained level. A granularity level \( l \) indicates that the analysis process produces only concepts at level \( l \) from the tree obtained from the Indiana Ontology by considering only the is-a relationship.

**Step 2:** Next, the rock carvings drawn in the area of interest are extracted from the repository and the associated ontology-driven metadata are clustered based on the information granularity level. In particular, a cluster is created for each concept \( c \) at level \( l \) in the ontology tree. Each metadata, which is a concept in the ontology, is associated with the cluster of the nearest concept at level \( l \). As an example, supposing that the concept \( \text{weapon} \) belongs to level \( l \), then the metadata \( \text{blade} \) is associated with the cluster \( \text{weapon} \). Notice that if the metadata belongs to a level lower than \( l \), then it could be associated with more than one cluster. Successively, each cluster is assigned a score computed by summing the inverse of the ontological distance between the metadata in the cluster and the concept \( c \) associated with the cluster itself.

**Step 3:** For each metadata in a rock carving, a 2D point is created based on the centroid of its cluster. The score obtained by the cluster is associated with the point. For each cluster, a layer is created by interpolating the points belonging to it; the higher the score, the darker the color of the layer. A polygonal summary layer is created by extracting the zones with highest score from each layer, and assigning them the color of their layer.

The Interpretation Summary View in Fig. 4 shows a polygonal summary layer. It visualizes on the map a polygon for each
cluster, and a pinpoint for their centroid. As an example, the places containing sacred carved rocks, e.g., priests and gods, are highlighted in semi-transparent blue. The archaeologist can infer that places in dark blue represent holy places for Mount Bego’s ancient people. The bar on the map allows the user to select the information granularity level.

Figure 4. Interpretation Summary View.

5.2 Discovering Interpretation Anomalies

The spatial relationships between carved objects represent precious information for the archaeologists. Indeed, archaeologists are interested in identifying correlations between the object spatial relationships and their interpretations. For instance, when an above spatial relationship holds between the mother goddess and the bull, then the interpretation is likely to contain the word birth (see Fig. 5). So far, even if with some criticisms by other archaeologists, most of these correlations have been identified in an empirical way, and have been used to solve the interpretation of complex or ambiguous engravings. By associating the relationships existing between the objects and the ontology metadata extracted from the interpretations with each engraving, it is possible to give evidence to the assumptions made so far by the archaeologists. However, the huge number of relationships and metadata contained in the repository suggests to develop a data visualization technique that can help archaeologists to examine, analyze, validate, and detect these correlations in a systematic way.

Figure 5. Four reliefs with the mother goddess above the bull [14].

The Metadata Graph View shown in Fig. 6 supports the archaeologist in finding new correlations starting from an ontology-driven metadata associated with the interpretations, such as birth. When the archaeologist selects a metadata m, the system extracts the set S of engravings whose interpretations have m associated with them. Each node in the graph corresponds to the interpretation of a carved object in the engravings in S, e.g., bull, while the edges connect the nodes for which there exists a spatial relationship within the engravings in S. The label associated with the edge indicates the type of the relationship. For instance, if an engraving contains a dagger overlapping the earth, then the nodes of dagger and earth are connected by an edge labeled with the overlap relationship. The size of each node is proportional to the frequency of the corresponding carved object in S. The color of the edge and the size of the label indicate the strength of the connection: if the spatial relationship indicated by the edge’s label occurs only in few engravings, no edge is drawn, but if it appears in many engravings, an increasingly thicker edge is used as the number of co-appearances rises. As an example, the graph in Fig. 6 shows that the above relationship between bull and mother goddess co-occurs in the engravings containing the word birth in their interpretations. This finding could suggest the archaeologist to look for engravings containing the bull, the mother goddess, and the above relationship between them, and whose associated interpretation does not contain the birth metadata. This search might help him/her discovery some anomaly in their previous interpretations. Let us indicate with S_i the set of engravings containing the bull, the mother goddess, and the above relationship between them. This investigation can be accomplished by inspecting the relationship through a double-click on the “above” edge between the bull and the mother goddess nodes. The Ontological View shown by the system presents detailed statistics on the semantic distance between the engravings in S_i. In particular, it visualizes a graph that highlights the ontological distance between the metadata m and the interpretation metadata associated with each engraving in S_i. The more distant from the center the interpretation metadata is, the higher the likelihood that the associated interpretation contains anomalies. Moreover, the darker the color of the nodes is, the higher the number of carved objects in the engraving is. As an example, Fig. 7 shows the ontological graph obtained by selecting the “above” edge between the bull and the mother goddess nodes. The graph highlights two interpretations whose metadata are semantically distant from “birth”. The one at the bottom of the figure contains few metadata in the interpretation; this indicates that the scene depicted in the corresponding engraving is not complex. The anomalous interpretations can be further inspected by clicking on the corresponding node: a window containing all the available information on the engraving is shown to the user, who can decide to revise his/her interpretation of the carving.

Figure 6. The Metadata graph shows carved objects occurring in the scene containing metadata “birth” in the interpretation, and the spatial relationships existing between them.
Figure 7. The Ontological graph highlights the ontological distance with respect to the “birth” metadata in the scene containing the bull, the mother goddess, and the above relationship between them.

5.3 Dating Rock Carvings

Dating rock carvings is one of the most complex tasks for archaeologists. The dating process is based on the comparison between the shape of the carved object, e.g., a weapon, and the shape of objects found during archaeological excavations on the site (see Fig. 8). In the case of Mount Bego, this is particularly challenging since the area was occupied during several phases of prehistory, and also during the recent history, until 1989 when the site was classified a historical monument and rock carving was banned. Thus, prehistoric rock engravings coexist with more modern carvings, made by shepherds, soldiers, visitors, and tourists from the roman period to nowadays [5].

Indiana Finder provides the Chronological Symbol View to support archaeologists in the dating process. Such view, as shown in Fig. 9, highlights the chronological distribution of carved objects similar to the one to be dated. Indeed, once the user selects a carved object $s$ interpreted as a symbol of type $t$ (e.g., a halberd), the system executes a shape-based comparison between $s$ and the objects interpreted as $t$ in the repository. The result of this process is summarized in a ring chart around $s$, which represents the chronological scale of Mount Bego’s engravings. The ring is divided into sectors, and each sector represents a prehistoric periods, e.g., the semi-circle from 4250 to 3550 B.C. indicates the Neolithic period. If the result of the comparison between $s$ and another carved object $x$ is above a given threshold, then it contributes positively to the period associated with $x$. We use a very high threshold of 0.9 since the shapes of carved objects of different periods often differ in little details, e.g., the shape of the blades. The size of the inner ring segments is determined by the number of objects of type $t$ in that period. Color indicates the number of positive comparisons in each period. A darker color means a larger number of positive comparisons.

This kind of visualization allows the archaeologist to perform a detailed analysis of which periods are more appropriate for a given carved object. When more than one period obtains the same highest score, the archaeologist can continue the investigation in different directions. First, s/he can analyze the carved objects that got the best results and compare them to object $s$. This is performed by moving the mouse over the time sectors. Second, s/he can analyze the co-occurrence of $s$ with another carved object $s_1$ of type $t_1$, positioned near $s$ in the rock carving. Once the user positions $s_1$ near $s$ in the view, the system computes the co-occurrence of the objects of type $t$ and $t_1$ in rock carvings, and updates the associated ring charts accordingly. In this case, the circle rotates in a way that the more appropriate period is the one that connects $s$ and $s_1$ (see Fig. 10). Finally, s/he can analyze the position of the carved objects more similar to $s$. This is performed by switching to another view, which visualizes on the map the distribution of the similar objects.

Figure 8. Carved weapons and examples found in prehistoric digs [5].

Figure 9. Chronological Symbol View of a carved weapon.
It is worth noting that the process of dating carving objects is strongly influenced by the discoveries in other archaeological sites. For instance, typological comparison with engravings in Valcamonica has been performed for Mount Bego’s engravings of the Neolithic period [5]. Thus, the importance of creating a single repository of the rock engravings spread all over the world is crucial for making the dating process more and more precise, and the work of archaeologists more and more integrated.

6. CONCLUSIONS AND FUTURE WORK
As stated by M.J. Morwood and C.E. Smith [29], “Rock art can provide socio-cultural information that is not generally available to the archaeologist. Despite this, it is only recently that rock art studies have been integrated into mainstream archaeology.” With the rapid growth of the interest in rock art archaeology, the need to support archaeologists in the inherently analytical nature of their investigations (become more and more complex due to massive increase in number and size of archaeological data), has rapidly grown as well. In this paper, we have presented a visual analytics software system, named Indiana Finder, which integrates computational, statistical, and semantic methods for exploring and understanding rock engravings. Indiana Finder can help archaeologists investigate complex patterns across spatial and temporal dimensions via clustering, image similarity, and visualization. In particular, it presents a set of views that highlight the relationships among rock carving reliefs, including semantic relationships based on metadata, and similarity relationships based on their shapes and on the surrounding symbols.

Current plans for improvements to Indiana Finder include the development of a multi-dimensional view for supporting archaeologists in the interpretation of new engravings, and the support of carved object motifs in the proposed views. Moreover, we intend to improve the visualization of chronological distribution of similar carved objects in the Chronological Symbol View by providing fine-grained statistics, such as considering the chronological distribution in centuries instead of age periods. Finally, we will investigate the use of query by sketch as a technique to ease user interaction and improve retrieval effectiveness in the repository by extending the agent-based framework proposed in [7].

7. ACKNOWLEDGMENTS
This research is supported by the “Indiana MAS and the Digital Preservation of Rock Carvings: A multi-agent system for drawing and natural language understanding aimed at preserving rock carvings” FIRB project funded by the Italian Ministry for Education, University and Research, under grant RBFR10PEIT.

8. REFERENCES