The 8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage
VAST (2007)
D. Arnold, F. Niccolucci, A. Chalmers (Editors)

Influencing Factors on the Visualisation of Archaeological Uncertainty

M. Sifniotis[1] P. Watten[1] K. Mania [1,2] and M. White[1]

[1] Department of Informatics, University of Sussex, UK
[2] Department of Electronic & Computer Engineering, Technical University of Crete, Greece

Abstract

This paper presents further work on a 3D visualisation system for the reconstruction of historical structures which takes into account archaeological uncertainty. The uncertainty associated with an archaeologist's interpretation is represented using possibility theory and visualised by shader-based information visualisation schemes. An increase or decrease in uncertainty is influenced by any related evidence recovered; we define this evidence as 'influencing factors'. Different types of archaeological evidence were identified after discussions with several archaeologists. To understand the individual importance of each influencing factor on an interpretation, we analysed data derived from formal questionnaires distributed to a selected group of archaeologists equally divided between Roman and non-Roman specialists. They were asked questions ranging from the wider perception of uncertainty to more specific ones on the identified types of archaeological evidence. We describe the stages involved in designing the questions, the process of gathering data and feedback from archaeologists, and the results themselves. Results suggest that specific evidence types are considered more favourably than others.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism J.2 [Physical Sciences and Engineering]: Archaeology

1. Introduction

During the last decade, the heritage community has witnessed an increase in three-dimensional (3D) reconstructions of archaeological structures. Assisted by advances in technology and cheaper technical costs, museums, archival institutions and heritage attractions have been increasingly using 3D reconstructions. However, archaeologists and computer scientists have urged caution in the abundant use of virtual reconstructions because of the possibility of misleading the public. Existing literature (see section 2) indicates the need for visualisations indicating where archaeological finds end and reconstruction begins.

Our methodology attempts to bring the element of the archaeological experts' uncertainty in the visualisation scheme. We define uncertainty as the extent of expert knowledge included in the visualisation of an archaeological structure. Archaeologists piece together available information derived from evidence into a speculative version of the past. This version becomes more certain as the evidence increases. We are using possibility theory in order to represent an expert's belief in his reconstruction, in combination with a number of factors and evidence that have been identified through questionnaires and structured interviews. The visualisation is implemented by using information visualisation schemes that have been extensively tested for perception. As a result, this approach offers visualisation levels that have as their basis the archaeologist's knowledge and information on reconstructed parts. For further analysis regarding this approach the reader is referred to [SJW*06].

One of the aspects of our work is to identify factors that may influence an interpretation, as well as how strong they are in their presence or absence. This paper presents related results from the interviews and questionnaires distributed to archaeologists. Section 2 looks at the existing literature related to archaeological uncertainty. Section 3 presents how the list of influencing factors were derived and the questionnaire was designed and distributed. Section 4 contains the data analysis and explanation of the results. Finally, Section 5 explores further work.

2. Background

Archaeology is both an uncertain discipline and a destructive process. An archaeological site can never be excavated and interpreted to its exact ancient dimensions; once it is excavated and its evidence removed and processed it cannot be repeated – it is therefore a one-way process. Consequently, archaeological hypotheses and interpretations also contain the element of uncertainty. As briefly mentioned in section 1, virtual reconstructions have only recently started acknowledging the issue. This has been introduced in work by Miller and Richards [MR94] as well as Ryan [Rya96]. A more formal initiative was attempted with the Cultural Virtual Reality Organization (CVRO) [FNRB00] and the publication of the Archaeology Data Service's guide to good practice in virtual reality. A more recent initiative is the London Charter, a set of "internationally-recognised principles for the use of three-dimensional visualisation by researchers, educators and cultural heritage organisations" [3DV07]. As a result, alternative approaches to virtual reconstructions have come forward, which usually involve differentiating between found and hypothesised parts, either by using dividing lines and colours [Eit98] sketch-based modelling [SMI99] transparency, colour and wireframe variations [ZCG05]. Furthermore, an approach to quantifying the reliability of a reconstruction has been introduced by [HN04]. A more analytical look at the related research can be found in [SMWW06].

3. Methodology

Our system [SJW*06] uses possibility theory to reflect an expert's belief on different parts of the interpretation. The system visualises the reconstruction with VRML/X3D and changes in belief are represented with ordinal pseudocolour sequences. For an extensive description of the theory and design of our system, the reader is referred to [SMWW06]. A Romano-British building known as Building III [MR05] near Fishbourne Roman Palace, UK is used as our case study. We have identified three categories through which uncertainty could be expressed:

1. Expert judgement: the archaeologist goes through each reconstructed part of the building and selects how sure they are about the interpretation from a list containing linguistic expressions of uncertainty.
2. Expert judgement with influencing factors: this second category asks the archaeologist to supply expert judgement and also identify any evidence which may influence judgement for the specific part.
3. Influencing factors: it would be interesting to observe whether by solely identifying available evidence one could estimate the uncertainty of each part.

The following section explains the concept of influencing factors.

3.1. Influencing factors

There is no set list of evidence types that turns up in every excavation. Additionally, every historical period may find an abundance of a specific type of evidence but a complete lack of another. We are interested specifically in factors that influence the interpretation of structures. In order to identify them, the first step was to conduct open structured interviews with archaeologists. Discussions were done on an individual basis and the panel consisted of five archaeologists. The discussion focused solely on the interpretation of structures. Three of the archaeologists were Romano British specialists, one focused on Eastern Europe and Byzantium, and the last on Anglo-Saxon England. All of them have in excess of 10 years experience in their fields. Three were interviewed personally, one by phone, and one by email. From these discussions a list of factors was created. Briefly this includes:

- Features: all elements of man-made structures; these can range from ditches to wall remains, to post-holes of wooden structures.
- Artefacts: any object made, affected, used, or modified in some way by human beings; this usually includes pottery, glass, lithic, etc.
- Biofacts: Biofacts or ecofacts constitute of human, animal and plant remains which are not changed by human interaction.
- Textual evidence: ancient texts and documents which may provide information on architecture, decoration, lifestyle etc of a specific culture.
- Absolute comparisons: a structure is compared to a structure of similar proportions.
- Contextual comparisons: a structure and its context is compared to similar structures with similar contexts.
- Topography: natural features of the landscape in which the building is located; elevation of the area, characteristics of the region and landform data in general.
- Peer Review: the interviews indicated that interpretation is heavily based on discussions and re-discussions with other archaeologists and architecture specialists were considered important.

The next step involved the creation of a questionnaire with an aim to gather more information on how archaeogists perceive these factors. The aim of the questionnaire was to determine if:

- There is any difference between the results obtained by Roman and non-Roman archaeologists. Our initial hypothesis was that a difference should be evident.
- There is any perceived preference among these different factors.

3.2. Questionnaire design

There is an abundant bibliography on the design of questionnaires for factors involving perception, immersion, and use of a system. However, in our case, there is no similar work conducted relating to questions and factors on archaeological uncertainty. As a result, the questions were designed progressively with consultation by the archaeologist panel described above. The archaeologists were asked to carefully read through questions and make suggestions on their validity, appropriateness, easiness of understanding, and order, as well as the types of answers available. A total of 35 related questions were created.

The questionnaire is divided in nine sections. The first section deals with perceptions of uncertainty in the discipline and how alternative hypotheses are handled. Sections two to six contain questions on the different identified factors. Section seven provides combinations of factors (for example features and artefacts, or artefacts and biofacts) and queries the expert as to how strong he/she considers this combination of evidence to be. In the eighth section, the list of factors is provided and the expert is requested to assign an order of importance for each. Finally, in the last section, the expert enters his/her personal data such as their specialisation field, time spent on excavations etc. Most importantly he/she is asked to give an opinion on the completeness of the factor list.

The answers to sections 1-6 are chosen from a seven point ordered response Likert scale [Lik55]. While prototyping the questionnaire, the first decision was to use a likelihood Likert scale, which would indicate the agreement or disagreement with a statement. However, feedback received from archaeologists showed that a frequency scale was better understood in accordance to the specific questions. As a result, a frequency scale was used.

Values include never, almost never, sometimes, often, very often, and always. The expert is asked as to how often he/she encounters each statement when making interpretations and reconstructions of a structure. The participants were aware under which group each question belonged since each section is clearly marked out as belonging to uncertainty, or a specific factor.

In section 7 the expert is asked to rate a list of factor combinations from 1 (weak) to 9 (strong). Lastly, section 8 follows a ranking system where the expert must rank the eight factors according to how important he/she believes they are in an interpretation. This is done by using numbers from 1 to 8 only once and assigning them to the factors.

Once the questions and scaling were completed, it was again tested by five archaeologists and five non-archaeologists for its coherence, layout and structure. The questionnaire was distributed in two versions, printed and electronic.

Both versions, printed and electronic, had the same content and in the same order. The printed version was personally distributed to archaeologists and collected after completion. The electronic version was created with a simple design using XHTML and PHP technology, while also being assistive and interactive. Features include error handling, data retention and real-time data collection. Before the actual distribution of the electronic questionnaire to any experts, fifteen people were asked to evaluate it in terms of: understanding of the instructions, text length readability, font size, questionnaire flow and error handling. They were also asked to identify their computer expertise, browser choice and screen resolution. 73.3% were male and 26.7 % female. 26.7% were less than 25 years old, 53.3% were between 25-33 and 20% were above 33. 33.3% considered themselves to be adequate or knowledgeable with IT while the rest were identified as experts or professionals. While 100% considered the length of text and instructions easy to read, 33.3% faced issues with the error handling and 20% with the default font size. About 73% of them were navigating with Mozilla Firefox and the rest with Internet Explorer. 60% had a resolution of 1024x768 and 40% had 1280x1024 or more.

Following their suggestions, we implemented more robust error handling, introduced the ability to change the text size, gave instructions for all popular browsers, and ensured full conformance with W3C XHTML/CSS specifications (Figure 1).

Figure 1: *Questionnaire page*

3.3. Participants

Participants were recruited across the archaeological discipline. A requirement for inclusion was the participation in archaeological excavation and interpretation of structures. Twenty experts participated, 50% specializing in Roman archaeology and the other half in different fields. The different fields excluded periods were limited structural and textual information is available - such as Early Ango Saxon England, Prehistoric, etc. As far as experience is concerned, 50% of the archaeologists had been working for 10 years or more on their particular field while 30% had also been excavating for more than 10 years. 60% of the archaeologists involved were between 26-33 years old, 15% between 34-41 and the rest above 41. 50% of the participants were male and 50% female.

3.4. Procedure

Before answering the questionnaire, the participants had to read an information page that explained the format of the questionnaire and how to answer questions. In the case of the printed questionnaire, this was also verbally explained. The archaeologists were advised that the questions were related to the reconstruction and interpretation of buildings and structures in general. Additionally the questions were applicable to cases where all available data has been collected from a site.

In the electronic version this was ensured by having to specifically tick a box in order to accept that they fully understood the instructions. At the end of the questionnaire participants answer questions about the questionnaire itself such as if they considered the factor list complete, or if the questionnaire was difficult to fill in.

4. Results

Our data was examined in three parts. The first part was related to the Likert-style questions (sections 1-6), the second part with the evidence combinations (section 7) and the last one with the ordering (section 8).

In order to establish the appropriate tests for our data, it was tested for parametric conformance assumptions. These assumptions presume that data: is normally distributed; has same variance through it; is measurable at least at the interval scale; is independent (results from one participant are independent from another). While the last two statements hold true, the first two were tested by normality and homogeneity tests respectively. Kolmogorov-Smirnov analysis indicated (sig. < 0.05) that our data is not normal, while Levene's test (sig. > 0.05) indicated homogeneity between the Roman and non-Roman groups. The violation of one assumption leads us to use non-parametric tests. The analytical results of these tests are attached in the appendix. The sections will be now described more analytically.

4.1. Sections 1-6

Thirty-five questions belong to these sections. Usually, one-way ANOVA is used to check for differences between independent groups (such as our case); however ANOVA requires parametric data. A non-parametric alternative to ANOVA is the Mann-Whitney approach. This test examines whether there are differences between the results of groups. The data was tested with exact significance values, an option which gives more accurate results in small samples like ours [FH02]. The Kolmogorov-Smirnov-Z (KGZ) test was chosen, which behaves similarly to the Mann-Whitney test but is much more appropriate for smaller samples. It appears that answers were not different between the Roman and the non-Roman group (sig. > 0.05).

The following step involved averaging the answers for each factor and examining the results. Terms that have a reverse phrasing were examined. For example, in the questions measuring the attitude towards uncertainty, the statement "There is only one true interpretation of a site" is a reverse-phrased one: if the attitude towards uncertainty is positive, the answer to this question should be negative. However, this would give a less weight in our scale, and vice versa. For this reason, six negative-phrased questions had their scores reversed before they were averaged. As a result, out of the 35 variables, 8 resulted (Table 1). This data was again tested

Resulting variable	#
Uncertainty	7
Absolute & Relative comparisons	3
Features	7
Topography	3
Artefacts	4
Ancient texts	3
Biofacts	3
Peer support	5
Total	**35**

Table 1: *Aggregated variables*

for parametric properties, and it was found to be homogeneous but non-parametric; Biofacts and Ancient texts failed the Kolmogorov-Smirnov parametric test (Table 2). Table 3 shows the results for the homogeneity test.

Roman experience		Kolm.-Smirn.		
		Statistic	df	Sig.
Uncertainty	yes	.255	10	.064
	no	.242	10	.099
Features	yes	.174	10	.200(*)
	no	.200	10	.200(*)
Artefacts	yes	.202	10	.200(*)
	no	.158	10	.200(*)
Biofacts	yes	.285	10	.021
	no	.184	10	.200(*)
Topography	yes	.185	10	.200(*)
	no	.214	10	.200(*)
Comparisons	yes	.138	10	.200(*)
	no	.178	10	.200(*)
Ancient Texts	yes	.302	10	.010
	no	.303	10	.010
Peer Support	yes	.202	10	.200(*)
	no	.139	10	.200(*)

Table 2: *Normality test for Sections 1-6*

A KGZ test was conducted to test for any difference between the two groups. The results, demonstrated in Table 4 shows that no difference was observed. The above results suggest that no difference can be observed between answers given from the Roman and non-Roman groups. Looking at the means of the different factors (Table 5 and Figure 2) we can see that archaeologists primarily place importance

	Levene Statistic	Sig.
Uncertainty	.491	.492
Features	.658	.428
Artefacts	1.590	.223
Biofacts	.179	.677
Topography	1.097	.309
Comparisons	1.300	.269
Ancient Texts	2.353	.142
Peer Support	.275	.606

Table 3: *Homogeneity test for Sections 1-6*

	Mean	Std. Dev.	Min.	Max.
Peer Support	5.2500	.75359	4.00	6.60
Comparisons	5.0000	.71737	3.67	6.67
Features	4.9625	.47659	4.25	6.00
Uncertainty	4.8429	.52006	4.29	6.00
Artefacts	4.7625	.66627	3.75	6.25
Topography	4.1500	.73727	2.67	5.67
Biofacts	3.9000	.72628	2.67	5.33
Ancient Texts	3.6333	.73270	2.33	5.00

Table 5: *Means for Sections 1-6*

	Kolm.-Smir. Z	**Exact Sig.**	**Difference**
Uncertainty	.447	.958	.200
Features	.671	.708	.300
Artefacts	.894	.332	.400
Biofacts	.671	.526	.300
Topography	.224	1.000	.100
Comparisons	.894	.181	.400
Ancient Texts	.894	.288	.400
Peer Support	.447	.981	.200

Table 4: *KGZ test for Sections 1-6*

on the suggestions of their peers (Peer support). Then, evidence from existing sites (Comparisons) and Features score quite strongly, followed by Artefacts and Topography, Biofacts and Ancient texts.

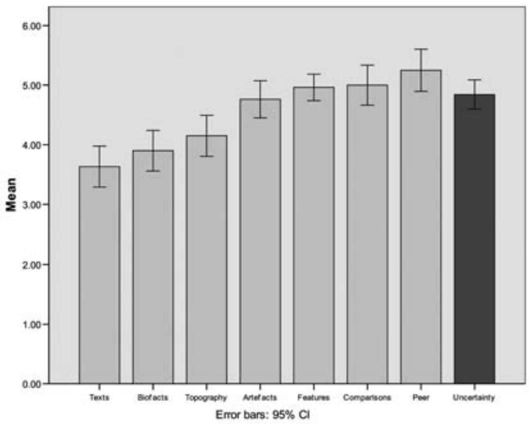

Figure 2: *Means of averaged variables*

4.2. Section 7

Twenty-two options belong to this section. Having established in Section 4 that the data is non-parametric, a KGZ test was conducted to test for any differences between the two groups. The results, included in the appendix, show no significant difference between groups and allow us to examine the means as a whole group. Table 6 shows the results from the combinations of section 7. It is interesting to note that combinations involving features, artefacts, contextual and absolute comparisons and topography score more strongly than other combinations. Additionally, combinations involving texts and biofacts score the least in the table.

	Sum	**Mean**	**Std. Dev.**	**Variance**
Features & artefacts	155	7.75	1.020	1.039
Features & contextual	138	6.90	1.071	1.147
Features & peer input	138	6.90	.852	.726
Features & absolute	137	6.85	1.424	2.029
Features & topography	131	6.55	1.638	2.682
Artefacts & contextual	130	6.50	1.821	3.316
Features & biofacts	129	6.45	1.638	2.682
Artefacts & absolute	128	6.40	1.635	2.674
Artefacts & biofacts	127	6.35	2.033	4.134
Artefacts & peer input	124	6.20	1.508	2.274
Topography & absolute	124	6.20	2.093	4.379
Topography & contextual	121	6.05	1.791	3.208
Artefacts & topography	121	6.05	2.038	4.155
Topography & peer input	118	5.90	1.586	2.516
Features & textual	115	5.75	1.970	3.882
Topography & textual	111	5.55	2.481	6.155
Biofacts & peer input	111	5.55	1.504	2.261
Artefacts & textual	111	5.55	2.114	4.471
Biofacts & contextual	109	5.45	2.328	5.418
Biofacts & absolute	106	5.30	2.319	5.379
Biofacts & topography	98	4.90	2.469	6.095
Biofacts & textual	98	4.90	1.997	3.989

Table 6: *Means for Section 7*

4.3. Section 8

The last section involves 8 variables, each accounting for a specific identified factor. The KGZ test (Table 7) indicated that there was no difference between the two groups (sig. > 0.05). The next step was to examine the group means. Table 8 and Figure 3 show the means calculations. Features and artefacts indicate higher scores for their means (6.85 and 5.70 out of 8 respectively) while the comparisons and peer input are between 4.50 and 4.90. Lastly, the biofacts and textual evidence are located at the lower end of the scale.

Factor	Kolm. Smir. Z	Exact Sig.	Difference
Features	.671	.498	.300
Artefacts	.447	902	.200
Biofacts	.447	.953	.200
Topography	.224	1.0	.100
Contextual	.447	.783	.200
Absolute	.447	.969	.200
Ancient Texts	.224	1.00	.100
Peer input	.447	.978	.200

Table 7: *KGZ test for Section 8*

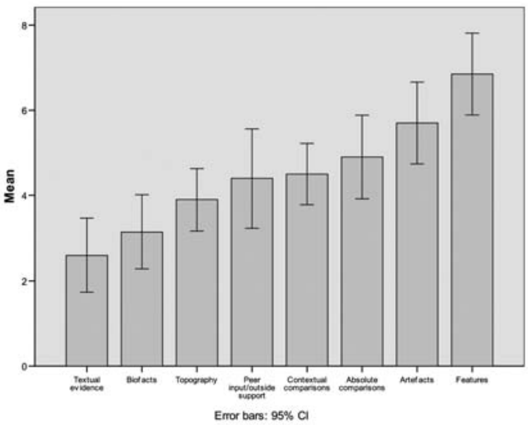

Figure 3: *Graphs of means for Section 8*

5. Conclusions and future work

The purpose of this paper is to present further work on a 3D visualisation system for the reconstruction of historical structures which takes into account archaeological uncertainty. We have analysed how the system involves the use of influencing factors; the archaeological evidence that is pieced together to create an interpretation. We have conducted a series of interviews and distributed questionnaires in order to understand how archaeologists perceive uncertainty and influencing factors. Because our first case study is based on a Romano-British structure, half of the participants were Roman archaeology experts.

Factors	Sum	Mean	Std. Dev.	Variance
Features	137	6.85	2.059	4.239
Artefacts	114	5.70	2.055	4.221
Absolute	98	4.90	2.100	4.411
Contextual	90	4.50	1.539	2.368
Peer input	88	4.40	2.479	6.147
Topography	78	3.90	1.553	2.411
Biofacts	63	3.15	1.843	3.397
Ancient Texts	52	2.60	1.847	3.411

Table 8: *Means of influencing factors from Section 8*

The questionnaire examined whether there is any difference between Roman archaeologists and those of other backgrounds when interpreting archaeological structures. Additionally, it examined whether there is any preferential order between the identified factors.

Throughout the three sections, results indicate that there is no significant difference between the two groups. Furthermore, the archaeologists seem to place more importance on peer feedback and evidence like features and artefacts. Accordingly, less importance seems to be placed to textual evidence and biofacts.

Feedback received from the evidence combination sections equally suggests that combinations of the *stronger* factors are considered more favourably when making interpretations.

It is worth noting that these results are not conclusive for the whole archaeological sector. Our participants were selected from closely *related* fields to Roman Britain, such as Ancient Greece and Byzantine. In cases where there is limited structural evidence available, such as Early Anglo Saxon England or prehistoric scenarios, the answers to these questions would have been different since we are solely dealing with reconstructions of structures. Additionally, Medieval times for example, give us an abundance of textual and historical evidence on which many archaeologists rely for their interpretation – this would perhaps reflect in a preference for textual evidence.

The results obtained show us that there is a preferential order which may alter our visualisation of uncertainty. For example, if an expert indicates a very high level of certainty for an interpretation based purely on ancient texts, then the overall level of confidence will be reduced due to the unreliability of the source. Conversely, if the interpretation was made on the basis of structural remains, then this would serve to strengthen the argument. In both scenarios, the resulting image from the visualisation stage would be different when using influencing factors when compared to using solely an expertŠs opinion. Software has been developed to incorporates these weights in a flexible manner which can be altered by an archaeologist focusing on a different time period.

In the longer term it would be fruitful for more studies like this to be conducted for different specialities. The combined opinions of various experts could be used as *profiles* for the interpretation of structures. The archaeologist would be able to choose between different weight schemes or add his own opinion to a specific period scheme.

References

[3DV07] 3DVISA: The London Charter, 2007. Accessed October 2007.

[Eit98] EITELJORG H.: Photorealistic visualizations may be too good. *CSA Newsletter XI*, 2 (1998), –.

[FH02] FIELD A., HOLE G.: *How to Design and Report Experiments*. ISBN 0761973834. Sage Publications, 2002.

[FNRB00] FRISCHER B., NICCOLUCCI F., RYAN N., BARCELÓ J.: From cvr to cvro: The past, present, and future of cultural virtual reality. In *Proceedings of VAST International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST '00* (Arezzo, Italy, 24-25 November 2000), Niccolucci F., (Ed.), BAR International Series 1075, Oxford, Archaeopress.

[HN04] HERMON S., NICCOLUCCI F.: A fuzzy logic approach to reliability in archaeological virtual reconstruction. In *Proceedings of the 2004 Computer Applications in Archaeology (CAA) Conference* (2004).

[Lik55] LIKERT R.: A technique for the measurement of attitudes. *Archives of Psychology 140* (1955), 55.

[MR94] MILLER P., RICHARDS J.: The good, the bad, and the downright misleading: archaeological adoption of computer visualisation. In *Proceedings of CAA Computer Applications in Archaeology CAA '04* (1994), Hugget J., Ryan N., (Eds.), pp. 19–22.

[MR05] MANLEY J., RUDKIN D.: *Facing the Palace: excavations just east of Fishbourne Roman Palace (Sussex, UK) 1995-1999,*. Sussex Archaeological Collections (SAC) 141, 2005.

[Rya96] RYAN N.: Computer based visualisation of the past: technical 'realism' and historical credibility. In *Imaging the Past*, Higgins T., Main P., Lang J., (Eds.). British Museum: Occasional Paper number 114, 1996, ch. Computer based visualisation of the past: technical 'realism' and historical credibility., pp. 95–108.

[SJW*06] SIFNIOTIS M., JACKSON B., WHITE M., MANIA K., WATTEN P.: Visualising uncertainty in archaeological reconstructions: a possibilistic approach. In *In ACM- SIGGRAPH 2006 Sketches and Applications* (Boston, MA, USA, August 2006 2006).

[SMI99] STROTHOTTE T., MASUCH M., ISENBERG T.: Visualizing knowledge about virtual reconstructions of ancient architecture. In *Proceedings of IEEE Computer Graphics International CGI '99* (1999), The Computer Graphics Society, IEEE Computer Society, pp. 36–43.

[SMWW06] SIFNIOTIS M., MANIA K., WATTEN P., WHITE M.: Presenting uncertainty in archaeological reconstructions using possibility theory and information visualisation schemes. In *In Proceedings of VAST 2006 7th International Symposium on Virtual Reality, Archaelogy and Cultural Heritage,* (2006), Ioannides M., Arnold D., Niccolucci F., Mania K., (Eds.).

[ZCG05] ZUK T., CARPENDALE M., GLANZMAN W.: Visualizing Temporal Uncertainty in 3D Virtual Reconstructions. In *Proceedings of the 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2005)* (2005), The Eurographics Association, pp. 99–106.