Doerr, Martin and Apostolis Sarris (eds) 2003. The Digital Heritage of Archaeology. CAA2002. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 30th CAA Conference, Heraklion, Crete, April 2002.

Matching ancient texts with geographical data

Lemonia Ragia and Edda Leopold

Fraunhofer Institut Autonome Intelligente Systeme
Schloss Birlinghoven, D-53754 Sankt Augustin, Germany
{lemonia.ragia,edda.leopold}@ais.fraunhofer.de

Abstract. Archeological sites are often mentioned with many geographical details in the texts of ancient authors. On the other hand aerial photography and remote sensing provide us with accurate geographical data. We propose a photogrammetric model which abstracts out the main features of the current geographical data. The ancient author's texts can also be reduced through text mining techniques to another description in the same model. We finally provide a methodology of matching patterns between the two descriptions which gives a clue on which current geographical site fits best the ancient author's description.
Key words : aerial photograph, text mining, modeling

1 Introduction

The merit of photogrammetry for the discovery, recording, mapping and documentation of archaeological sites has long been recognised (Green 1983). Archaeologists often experience the problem of not being able to comprehensively separate the abundance of ancient textual sources which might be related to an archaeological site. In archaeology it is not a simple process to find the correct places where the excavation has the best chances. There are sometimes many different sources of texts in different languages and it is an exhausting problem to find them, to translate them and to understand them. It takes time to search in order to find all available information. It is a very expensive process to excavate without knowing where the locations are. If archaeologists find many texts they have to study the results available and formulate a hypothesis. Then they have to find evidence that can support this hypothesis. What is really important is the final decision and the location of the place where the excavation can be started. Despite the search in the texts it is possible to have problems to find the right places because either the description of the texts is not good enough or the places have been changed in the pass of time. It can be that other constructions have been built on the site. In these situations it is very helpful to have a big overview of the scene.

In this work we propose a method for making a hypothesis where an excavation can be started more reliably. Our modeling approach is based on two other models: the first model is taken from text mining and it is *text based* and the second one is created from the photographs and it is *image based*.

With an *image based* model based on photographs one can reconstruct the 3D Model of the scene with the position and the form of the objects. In addition to the geometric reconstruction of the objects we can also have the orientation of the objects, the distance between them and the topological relations. All this information can be obtained without physically having access to the site. We can obtain these information from images taken from plane or satellite, which allow a large scale overview of the scene.

It is not a trivial task to compare texts and image – based models.

Texts and images constitute different semiotic systems in which information is conveyed in a completely different way. In text, objects are referred to by words, and the relation between objects is expressed by grammatical structures in the text. The meaning of words is not unique – one word can refer to different classes of objects - , and the same object can be referred to by different words or even by references within the text.

Recently there has been intensive research on the automatic semantic annotation of images (Barnard et al. 2002). A broad range of computer vision methods have been used to search collections of images. Typically images are matched based on features computed from the entire image or from image regions. A review of these techniques can be found in (Forsyth 1999 and Barnard et al. 2002). An example is the task of predicting words associated with whole images (auto-annotation) and corresponding to particular image regions (region naming). Region naming is a method of object recognition as a process of translating image regions to words, much as one might translate from one language to another.

Although Barnard et al. find relations between images and words their approach is not useful for our purposes. Our task is not just to analyze a picture into regions and translate the regions into words, but finding correspondences between geographical and textual information. The 2D model from the images is not arbitrary pictures, but well defined carriers of information. Texts are not only a sequence of words with a well defined meaning, they also establish fine grained structures among the words and the objects they refer to. These structures have to be analyzed and geographic information has to be extracted in order to match information from images and texts.

2 Text Mining

Text mining is the process of looking for interesting and useful patterns in natural languages texts and to search for interesting relationships between the extracted entities (Fig. 1). In (Fayyad et al. 1996) text mining is defined as identifying textual patterns which are "valid, novel, potentially useful and ultimately understandable". Translated to the task of matching texts with geographical data this means: identifying those textual patterns which can be related to information in a geographical image. Text mining can extract key concepts from texts, organize documents according to a given schema and discover themes in a collection of documents. There are four basic steps:

- Information retrieval : a pattern of interest keywords are used, in order to find relevant documents within a large set of possible relevant documents.
- Information extraction : interesting information for a concept is extracted from the selected documents.
- Information mining : data mining techniques are used in this step in order to discover patterns within data.
- Interpretation : the last step is to place an interpretation on the patters extracted from the mining phase.

Text classification is one of the basic techniques in the area of text mining. It means that text documents are filtered into a set of content categories. For the task of text classification, there exist promising approaches, which stand for different learning paradigms. Support vector machineds (SVM) are one of the most successful solutions (Joachims 1998). The training of a SVM requires many positive and negative examples for each class to be considered.

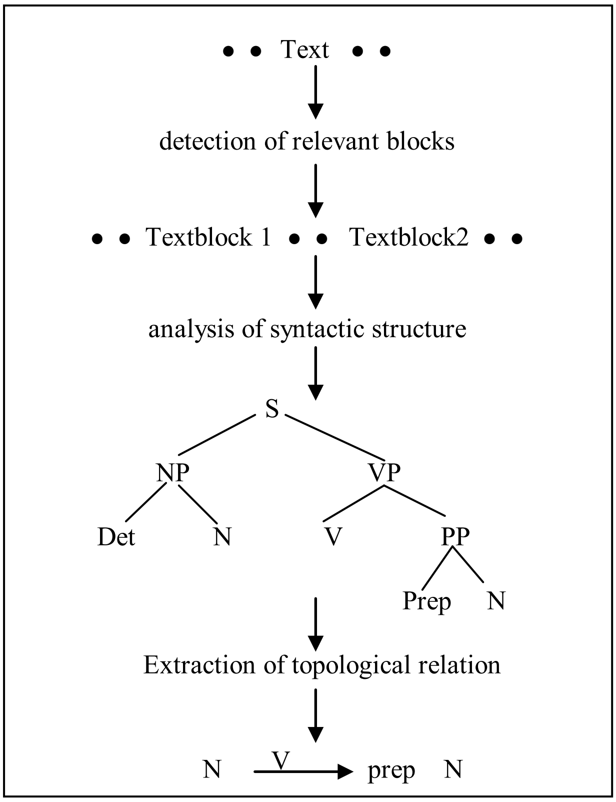


Fig. 1. Shows the different steps of text mining from a large amount of documents.

Other techniques for finding different groups in the documents are clustering and categorization. The clustering algorithms divide the documents into similar in content groups. The topics or themes for clustering the documents are found automatically and not defined by the users. Categorization techniques enable the system to assign documents to categories predefined by a taxonomy.

There are many techniques for information extraction (Grishman 1997). The information extraction systems have in many cases syntactic analysis and learned patterns based on pure word sequences (using regular expressions and Markov models) (Grishman 2001). Mining the information is based upon data mining techniques (Hand et al. 2001). Finally, interpretation requires much expertise in the application area, in this case archaeology.

3 Photogrammetry for archaeology

A great contribution to archaeological fieldwork and recording has undoubtedly been made by aerial photography (Greene 1983). With aerial photography one can see many details of sites. The visual effectiveness of photographs has been appreciated since 1850's. A ground observer is too close to the objects and has difficulties to see a pattern. The comprehensive view from the air can show significant features and relationships between the objects. The photograph is sometimes the only clue that something is really there. There are four basic characteristics which are important for archaeologists (Greene 1983):

1. The *illustration*. It is a simple but important function. Sometimes there are many difficulties to see things from the ground. For example there is no trace in some places of objects on earth or there are soil marks or crop marks which one can only see from above. An overall view of an air image avoids these difficulties.
2. The *search*. This is especially valuable in order to study the mutual relationships on the surface, to discover the results of the investigations and to analyse the different crop mark sites.
3. The contribution of images to *excavation* and *fieldwork*. The excavation can be undertaken because of its key position in the development of the local landscape or because the special site is a good fit from the study of the images.
4. The images play a basic role of *conservation*. The information can be stored and it is the evidence for the different phases of excavation on the archaeological site.

The great advantage of the photography is the overview of the place. There are also a lot of other advantages:

1. the objects can be documented without touching them,

2. the objects and the sites can be reconstructed in 2D/3D model,
3. all the features can be checked and documented at any time,
4. the accuracy and homogeneity of the photogrammetric analysis,
5. the whole excavation can be documented and this document can be standardized,
6. the information can be put in CAD systems and it allows computer rendering and animation and
7. with the GIS Systems a large number and variety of analysis tools are available.

There are two different types of images: the vertical view and the oblique view. In the first type each image should be taken directly above the subject and in the second type the images can be taken at shallow angles of about 30° to 45° from the horizontal. Sometimes the angle can be 50° or more from the horizontal, in order to reduce perspective distortion. The use of these types depend on what has to be excavated and studied.

The scale of the aerial images varies between 1/50000 and 1/1000. The scales from 1/50000 until 1/20000 are used for natural features and very large man made objects. The scales of 1/10000 and smaller are used for small areas and for small constructions. The satellite images are produced in very large scales. They have a very good resolution and they can provide an enormous amount of useful information. One can also use thermal images or airborne radar images.

In digital photogrammetry there are many techniques which contribute to a better photogrammetric analysis with high accuracy and efficiency. There are many image processing techniques. Some tasks of these techniques are the visualisation of images, image enhancement, information extraction (Weidner 1995, Faber 1998). The extraction of the third dimension plays a great role (Förstner 1995). Especially the extraction of the digital elevation model DEM (Krzystek 1998) can be used for the detection of break lines and significant points for an excavation. All of these aspects can be combined in orthophotos which are rectified photographs that have a unique scale.

Photogrammetry also provides useful methods for object extraction (Fuchs et al.1998). The objects can be extracted not only with the classical techniques but also with new semi-automatic procedures (Förstner 1999, Gülch et al. 1998). In our approach we discuss a methodology which uses the topological relations between the objects (Ragia et al. 1998). The benefit of this approach is that the photograph is translated into a format which is more compatible with the characteristics of textual information.

The input for object extraction is one image for a 2D model or two images for a 3D model of the same scene. The following steps are required:

- The objects which are shown in the images can be extracted in a stereo plotter, or in a semi automatic system for object extraction.
- For the analysis of the topological relations the model of the Egenhofer is used (Egenhofer et al. 1997). The topological relations {*disjoint, touch, overlap, covers, covered by, contains, contained by, equal*} are calculated.

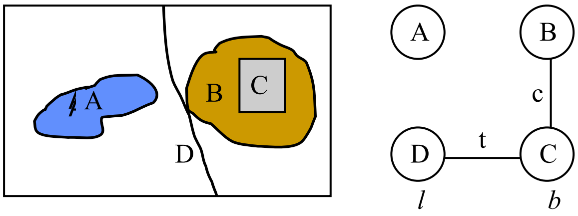


Fig. 2. Four objects with their *feature adjacency graph*.

The *Feature Adjacency Graph* (Fuchs et. al. 1995) (Fig. 1 on the right) is used in order to show the objects with their topological relations. This graph is a bibartite graph with attributes. The nodes of the graphs represent the objects of the image. The attributes of the nodes refer to the objects. The attributes of the edges indicate the topological relations.

Consider the example in Figure 2 with four objects: On the left there is the object extraction and on the right the corresponding *feature adjacency* graph is shown. The letters A, B, C, D indicate the objects. The letter *t* means *touch* and *c* means *contains*, the *b* indicates a *blob* (surface) and *l* a *linear object*.

4 Images and texts

Whereas the interpretation of written text belongs exclusively to typological semiosis, *image models,* symbols themselves, belong, nevertheless, to a small set of types, which is uniquely defined in the legend and make use of both topological and typological semiosis. The vocabulary of a language is discrete: different words have different meanings and there is no gradual change from one word to another in contrast to locations on an *image model*. Symbols can be more or less distant and angles can also vary by degree. The symbols themselves, however, belong to a small set of types, which is uniquely defined in the legend.

	Image model	*Text model*
syntactic relations	topologically defined	defined by grammatical relations
lexical information	uniquely defined in the legend	ambiguous and context sensitive
orientation	precise and continuous	fuzzy and discrete
distance	precise and continuous	fuzzy and discrete
interpretation	objective	subjective
thematic	discrete	Discrete

When an aerial photograph is transformed into an image based model using photogrammetic techniques, the representation of the information changes. Figure 2 displays a simplified 2D *image model* on the left hand side. The positions of the objects on the 2D image model are transformed to adjacencies in the *feature adjacency graph* on the right side. In the 2D image model quantitative differences of degree in the position of an object lead to a difference of degree in its meaning ('topological semiosis'). The *feature adjacency graph* on right side constitutes typological information, because two objects are adjacent or not, there is no degree in adjacency ('typological semiosis'). This photogrammetric transformation makes the information contained in the 2D image model more compatible to textual information, since texts make exclusively use of typological semiosis

Texts constitute a hierarchy of linguistic units. Written texts are made up of sentences, which in turn consist of clauses, words, morphemes and graphemes. Real world objects are usually referred to by words (lexical semantics) whereas the relation between the objects is conveyed by the syntactic relationships between the words within a sentence (propositional semantics).

The problem with lexical information is, that it is by no means unique. Apart from technical terminologies the meaning of a word is not uniquely defined in a lexicon, it depends on the interpretation of the reader as well as on textual context. This is why (Rieger 1989) introduced the notion of fuzzy semantics. Fuzzyness of linguistics categories enables us to describe a continuous reality with a discrete set of linguistic entities. Furthermore a lot of words are polysemious, which means that they have various different (fuzzy) meanings. The English word *bank* for instance can refer to a park bench or to a financial institution. On the other hand words usually have synonyms, which can refer to the same class of objects.

What is important for the problem that we consider here is that especially sizes and distances are mostly expressed in a fuzzy way (Tab. 1). The expressions *near by*, *not far away from*, or *distant from* do not correspond to an exact distance of objects. They can be interpreted as referring to a fuzzy distance. This means for each quantifier *A* there is a membership functions which assigns to each distance truth value $\mu_A(x)$ which indicates how much the proposition *A means x* is true. Truth-values range between 0 (=false) and 1 (=true).

Words or phrases in a text often refer to other words which have occurred previously in the text. In the example *The bridge is made of stone. It crosses the Rhine river.* The word *it* refers to the bridge which was mentioned in the preceding sentence. (Grosz et al 1989) gives an overview of different methods to evaluate these referring expressions.

5 Matching the two models

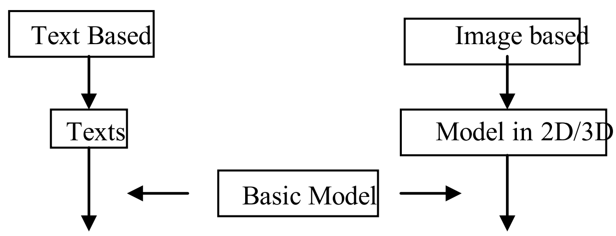
Table 1. Representation of different parts of information in *image* and *texts models*.

The *image model* as described in Section 3 provides a basic model of a scene. The scene can have additional geometric details in which we are not interested and are abstracted out. Actually we are not interested in the minutest details. In addition there are new objects represented in the images which did not exist originally at the archaeological site. Texts on the other hand are time references of the site and represent pertinent information of objects as they were. We have to be careful in order to have better results with our approach:

- We must distinguish between natural and man made objects. Natural objects do not have too many changes over the years.
- We must distinguish between the 'new' man made objects and the 'old' original objects of the site.

The output of the *text model* is a pattern with keywords and prepositions as in Figure 1. The output of the *image model* is a graph as in Figure 2. For the scene there can be many interesting patterns in the text documents and many images in the image model. The question is how to find the interesting patterns with the corresponding image. All combinations must be examined.

Consider as an example the scene of the Figure 2. Assume that we have found the following text: '..The grave is situated not far away from the lake. The native people live in the castle at the top of the hill. Every day when the people go down to the lake to fetch water they pass the grave..' The keywords in this text are hill, grave, lake, castle. After the text analysis there are two spatial relations between grave and lake, and castle and hill. We have the graph in the Figure 2 on the right of the *image model*. Then you can check the correspondence of the objects of the graph with the objects in the output of the text based model. In parallel you can check the spatial relations. The basic model is based on the correspondences of the objects with high reliability (Fig. 3). The object A is without doubt the lake in the scene, the object B is the mountain and the object C is the castle. The object D does not exist in our textual description, but it could have been built later. One spatial relation is matched between castle and mountain. The other one gives us the clue where the grave can exist.



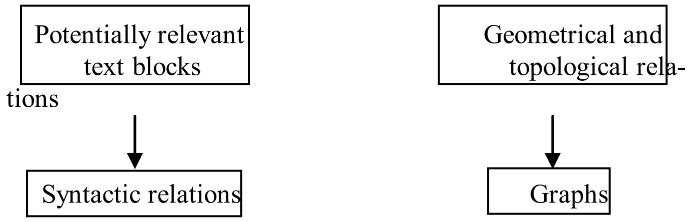


Fig. 3. The matching process between the two different models. In the left part of the figure there is the main steps for the text based model and the right part shows the main steps of the image based model.

6 Conclusions

Text and images are different semiotic systems. It is very difficult to combine the two different technologies. Most information of an image model is precise in contrast to the text model where the information is in a fuzzy way. Both models are separately used for archaeology. Photogrammetry is being used for many years in archaeology but text mining is quite a new method. Both texts and images can be handled in an automatic way, making large amount of data accessible. There are many techniques for extracting information from the *image* and *text model* with high reliability.

We have shown that there are synergetic effects which help to combine the two different techniques and result in better information for the archaeologists. Up to now this approach is on a theoretical level. Empirical results can contribute to the application of this approach. Users of this approach could have two main advantages: a) they have spent not too much time in order to look for the information and b) they need not read all these documents.

References

BARNARD, K., DUYGULU P., FORSYTH D., DE FREITAS N., BLEI, D. and JORDAN M. I., 2002. Matching Words and Pictures, submitted for publication in Journal of Machine Learning Research.

EGENHOFER, M. J, FLEWELLING, D.M. and GOYAL, R. K., 1997. Assessment of scene similarity. Technical report, University of Maine, Department of Spatial Information Science and Engineering.

FABER, P., 1998. Digital Image Processing, Proc. Third Course in Digital Photogrammetry.

FAYYAD, U. M., PIATETSKY-SHAPIRO G. and SMYTH P., 1996. From Data Mining to Knowledge Discovery: An Overview. In: Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy R., (eds.), *Advances in Knowledge Discovery and Data Mining*. Menlo Park, California: AIII Press/MIT Press, 1-34.

FÖRSTNER, W., 1995. GIS – The Third Dimension, Proc. IUSM WG on GIS/LIS Workshop “Current Status and Challenges of Geoinformation Systems”, Hannover, Germany.

FÖRSTNER, W., 1999. 3D-City Models: Automatic and Semiautomatic Acquisition Methods, Photogrammetric week, Stuttgart.

FORSYTH, D., 1999. Computer Vision Tools for Finding Images and Video Sequences, in: Library Trends, 48 (2), 326-355.

FUCHS, C. and FÖRSTNER, W., 1995. Polymorphic Grouping for Image Segmentation, In 5th ICCV, Boston, pp. 175-182.

FUCHS, C. and HEUEL, S., 1998. Feature Extraction, Proc. Third Course in Digital Photogrammetry.

GREENE, K., 1983. Archaeology, An Introduction.

GRISHMAN, R., 1997. Information extraction: Techniques and challenges. Information Extraction: Amultidisciplinary Approach to an Emergine Information Technology, Vol. 1299: 10-27.

GRISHMAN, R., 2001. Adaptive Information Extraction and Sublanguage Analysis, in Proc. of Workshop on Adaptive Extraction and Mining at 17th International Joint Conference on Artificial Intelligence, Seattle, USA.

GROSZ, B. J., POLLACK M. E. and SIDNER, C. L., 1989. Discourse; in: M. Posner (ed.), Foundations of Cognitive Science, London: Bradford MIT Press, pp. 437-468.

GÜLCH, E., MÜLLER, H., LÄBE, T. and RAGIA L., 1998. On the performance of semi-automatic building extraction, Proceedings of ISPRS Commission III Symposium, Columbus, Ohio, 1998.

HAND, D., MANNILA, H. and SMYTH P., 2001. Principles of Data Mining.

JOACHIMS, T., 1998. Text Categorization with Support Vector Machines: Learning with many relevant features. European Conference on Machine Learning (ECML), Claire Ndellec and Cline Rouveirol (ed.).

KRZYSTEK, P., 1998. Automatic Aerial Triangulation Conceprual aspects and new investigations, Proc. Third Course in Digital Photogrammetry.

RAGIA, L. and WINTER, S., 1998. Contributions to a Quality Description of Areal Objects in Spatial Data Sets, Proceedings of ISPRS Commission IV Symposium, Stuttgart, Germany September 7-10.

RIEGER, B. B., 1989: Unscharfe Semantik; Lang P., : Frankfurt a.M.

WEIDNER, U., 1995: Image Processing Techniques, Proc. Second Course in Digital Photogrammetry.