Geoinformatics and Crowdsourcing in Cultural Heritage: A Tool for Managing Historical Archives

Klára Paclíková1, Anne Weinfurtner2, Marta Vohnoutová1, Wolfgang Dorner2, Jan Fesl1, Michal Preusz1, Libor Dostálek1, Kateřina Horníčková1

1 Institute of Applied Informatics, Faculty of Science, University of South Bohemia České Budějovice, Czech Republic
2 Deggendorf Institute of Technology, Institute of Applied Informatics, Germany

Abstract

Archives of historical photographs have a great potential for "geo- or spatial sciences", for they can provide highly relevant visual data on historical landscapes, populated places and settlement structures, including those now destroyed. Processing of these archives represents many challenges, among them the application of geoinformatic concepts and information technologies. The article presents the example of geo-referencing, crowdsourcing, and other computer-based technologies applied to the archival photographs of today-destroyed sites on the Czech – Bavarian border, where many villages, farm sites and monuments were destroyed in the 1950s or abandoned as a consequence of post-WWII development. In the situation of dramatically changing landscape and land use, historical photographs are an important source of documentation for both research and virtual reconstruction of disappeared places, landscape, and society.

Keywords

Geo-referencing, historical photographs, crowdsourcing, landscape reconstruction, cultural heritage, image analysis, digital humanities, digital archive, IT methods in cultural heritage, 3D reconstruction, destroyed cultural heritage.

Introduction

Landscape and land use in the Czech Republic have changed significantly during the past hundred years, i.e. in the "contemporary" or "recent past" (Harrison and Schofield, 2010; Holtorf and Piccini eds., 2011; Krajic et al., 2017; Vařeka, 2013). Turbulent social and historical development associated with the outcomes of World War I and II, such as the expulsion of the German population (Ahonen, 2010; Arburg, 2005; Glassheim, 2016; Staněk, 1991; Vařeka et al., 2008), settlement of newcomers (Pešek, 1986a; Pešek, 1986b; Pešek, 1986c; Pešek, 1986d), Socialist regime’s collectivization (Blážek and Kubálek, 2008) and closing of the border and the construction of the Iron Curtain (Jilek and Jilková, 2006; McWilliams, 2013), transformed human lives, local social structure, and, especially in the border regions, the rural landscape defined by economic and agricultural production.

Important witnesses of these dynamic processes are historical photographs in the archival collections that reflect cultural, social and economic developments, nature, and climate changes, and provide visual sources for the interpretation of past events, activities and situations. Moreover, historical photographs are emotional documents and media of memory of the past that re-enact historical situations in the photographic medium. Besides the existence of specialized photographic archives, regional archives and museums hold an increasing number of photographs, images and digitized visual material documenting past situations. Preserving and presenting this rich material is one of the most important contemporary challenges in archival science. Interpreting historical photographs needs research environment, in which they are embedded, consisting of technical and content information (meta-data) as well as references to spatial and temporal contexts in which they were taken, and other relevant textual data, documentation and retrieval functionalities.
This means that the digitisation of archives is not the only task today, but also the improvement of documentation is needed to make archival material accessible and usable for research.

The aim of the article is to present the interdisciplinary approach and the first results of the Czech-Bavarian project PhotoStruk, focusing on geo-informatics and crowdsourcing as computer-based methods to speed up the process of classification and meta-data documentation of archives of historical photographs. Together with methods of non-destructive archaeology and history, they support research into the reconstruction of abandoned settlements, and lost historical landscape, and cultural monuments of the Šumava (Böhmerwald) border region. The region on the south-eastern border of the Czech Republic suffered from the abrupt social and economic changes as a consequence of the post-WWII development: expulsion of Germans and only partial re-settlement by Czechs, fundamental change in land use due to the reforestation, creation of military zone and destruction of numerous villages, farms, and sites. The IT method and tools are tested on a collection of photographs from the archive of the Museum Fotoateliér Seidel in Český Krumlov, Czech Republic. A family-held photographic atelier archive, it contains 140 000 photographs taken from the late 19th century until the 1950s. Historical photographs of Fotoateliér Seidel have not yet been systematically investigated in the research of extinct villages in the mountainous region of Šumava. Its digitized collection contains images of today-destroyed or abandoned villages in the region that provide valuable information for the virtual reconstruction of the sites, but also a number of undocumented and wrongly-tagged images that will be newly reassessed in the framework of the project, and valorized for their information value.

### Material and methods

Photographs provide a rich source for identification of spatial changes over a period of time, such as development of historical landscape situations, settlements and buildings. Disciplines such as monuments protection, landscape planning, as well as geomorphology, history and archaeology widely use them (Turner, 1990, Kadmon and Harari-Kremer, 1999, Chandler and Brundstén, 1995, Yilmaz et al., 2007, Arias et al., 2006, Estes and Harari-Kremer, 1999, Chandler and Brundstén, 1995). Tagging of books with a geo-reference to provide spatial metadata is an established approach in archival sciences (e.g. Lu et al., 2010). In terms of computer-based crowdsourcing, first attempts using it to interpret historical photographs have already been made. Kalfatovic et al. (2009) published images from the archive collection on the photographic platform Flickr to study public interaction with historical images, and to explore how this can help improving meta documentation. The project generated a huge public interest and a large number of data were provided by the public. Due to the technical limitation of the Flickr platform, user interaction was restricted to writing comments on the photographs, which had to be analysed manually for the extraction of semantic and spatial information. Crowdsourcing and geo-referencing was also applied to historical maps (Bill et al., 2015), indicating that the public can contribute to handle large amounts of archival data.

From the perspective of geomatics, the main
The difference between geo-referencing of maps by Bill et al. (2015) and Kalfatovic et al. (2009) is the level of accuracy, and the concept of sourcing geospatial information. Collecting geonames from comments, as applied by Kalfatovic et al. (2009), can provide spatially vague results, because geonames refer to either smaller or larger areas (theoretically ranging from field names to the names of districts or states). Working with coordinates for geo-referencing of maps, as used by Bill et al. (2015), requires skills in the interpretation of maps and forces the user to decide on a particular location (identified by the coordinates), where the user may possess only a vague or imprecise geospatial knowledge. His or her contribution can still be helpful, if he or she can provide at least some limited information, e.g. geonames in a comment.

The combination of both concepts may (1) be used to make crowdsourcing system accessible for a broader audience without map-reading skills, and can (2) be used to improve step by step the quality of geospatial information by narrowing down geonames to geo-coordinates involving information by several different users. To overcome technical and legal problems in using established platforms such as Flickr, the PhotoStruk team has developed own web application that integrates comments and geonames as well as geo-referencing based on maps and coordinates to provide a suitable data for 3D reconstructions of lost buildings and villages (Remondino and Campana, 2014).

**Project rationale**

Our project develops tools, which combine computer-based technologies to better extraction of valuable spatial and content metadata from the archive stocks (Figure 1). The project is based on the assumption that by means of automated image analysis, geo-referencing and crowdsourcing, a faster and more targeted access to the photographic archive material can be provided for the public, and a number of new metadata acquired, or improved in return. The project regards crowdsourcing as the computing method capable of mediating the knowledge.
of regional experts, local, and informed people (Košátková, 2011).

The project combines algorithmic approaches with controlled user interaction to generate documentation to historical photographs, and open up archives to the public interest. To achieve its goal, digitized photographs from an archive are made accessible through a specially designed web platform. This allows the interested public to access to the archived images and data, and invite the interaction in the form of comments and annotations, locating and geo-referencing on the map or providing other types of data, e.g. (historical) place names, identifications or texts. The crowdsourcing tool thus gives the broad public the opportunity to participate in research and identify displayed objects, locate them on a map and set up the photographer’s point of view (Figure 2).

- **Batch inputs**
  Partners of the PhotoStruk project (e.g. Museum Fotoatelier Seidel) are the main sources of data, which can be sent in batches. Data can also be digitized from non-digital sources.

- **Archive**
  The core of PhotoStruk is a long-term archive, which contains all the data, and its descriptions. Only authorized archivists and specialists can access this archive. They verify and complete the data and decide, which data can be made accessible to the public via the web portal.

- **3D modelling**
  The projects also aims to analyse existing and newly-generated metadata, and use them for 3D modelling of lost monuments and building structures. 3D models of historical objects will be made accessible to the public via the web portal.

- **Webportal**
  The web portal is the central platform accessible to the public. Everyone can not only view the historical data, 3D models, images, and maps, but also add other data or images, and complete or correct the existing ones. This facility is based on the assumption that this contribution by public (known as citizen science) will be a valuable asset to the meta data documentation. Except the public, other external resources, e.g. Wikipedia and Geonames (geographical database), can be used to draw information from.

- **Map**
  The interactive map, and the facility to locate data and images on the map is the most essential part of PhotoStruk. The map has an additional dimension - time. Users can travel in time back and forth, and see the countryside of the specific periods, data related to the period, and photographs of that time.

- **Data exchange**
  The data sets of the archive and web portal are managed in two standalone databases, which have no direct access to each other. A controlled data exchange process allows data transfer without overriding any data sets. The database of the web portal is designed for geoinformation system operations. After verification of spatial and semantic metadata, the information is stored in the archive.

- **Application for archivists and specialists**
  The only people who have access to the archive are archivists and authorised specialists, who have to get authorization before accessing the archive. They work with the archive data with the help of the application created directly for them (Figure 3).
The web portal interface consists of the following main screens:

- Main webpage provides the access to system functionalities
- Fast villages overview - shows the basic information about all villages (i.e. sites) in the database
- GPS coordinates selection dialogue window allows to outline (by drawing contours) the specific geographic area of interest
- A photograph is assigned to a specific location (village) on the map, which defines a relation between a specific place and image (see the Figure 2)
- Interactive image searching tool is able to find the most relevant images according to the given keywords and sorts the results by the proximity value.

In the process, different ways of adding data to an image or its location on the map are offered to the users. Users can locate photographer’s position in the map directly by marking a point on a map; mark a known object in the photograph, whose location can be drawn from external data sources; give a geoname from the keywords or identify it in the published comments. Objects or landscape features on unknown photographs could also be compared to other images with similar features that have already been located, and this way an unknown photograph could be identified. This process can be repeated infinitely, which allows for multiple cross-checking between geo-referencing – this way crowdsourcing can contribute new data to a given image or data set. Upgrading of the data sets will be achieved by continuing process of multiple suggestions by different users, which improves the accuracy in successive steps.

The project develops in eight work phases (Figure 4), with the first step being the data collection and analysis. The data contains photographs, postcards and other digitized documents, which provide visual or textual information about the site. The processed data collection is published through a web portal for the interested public, e.g. local citizens, family members and descendants, who remember pre-war situation and the destroyed settlements, amateur historians, researchers, and students. The web portal is the central platform for interactive testing of of geo-referencing tools, web presentation, and direct users’ interaction with digitized historical photographs. We are interested in the ways how public interacts with the images, and how substantial interest can be generated and used in the processing of the photographic archives in a targeted manner.

The rationale behind using crowdsourcing method is the assumption that one or a few persons would not be able to manage such a large data volume as a whole archive, and may be lacking local knowledge or expertise on a required geographical scale. Thus the images will be first analysed by automated image analysis technology, and then tested in crowdsourcing. The project also tests how to verify the crowdsourced data based on interactive user questioning. Thus it will be able to provide first Central European data on the usability of crowdsourcing for research in the humanities, and meeting the standards
of scientific work with the aim to eventually be integrated into the classical methods of archival work.

**Identification, documentation and interpretation of landscape changes**

Once historical photos are digitized, registered in database and georeferenced, it is possible to supplement them with other information, drawn from non-destructive archaeology and history research, as shown on the example of Kapličky (Kapelln) village.

One of most important sources of added (mainly topographic and landscape) data is remote sensing, i.e. Light Detection and Ranging (LiDAR) and Airborn Laser Scanning (ALS) methods, which create the Digital Terrain Model (DTM). Relief map image DTM obtained from the Czech Office of Surveying and Cadastre (ČÚZK), which was processed in the Geographical Information Systems (Figure 1), was used to identify anthropogenic relics preserved under the vegetation (Gojda et al., 2013). The Lidar data were correlated with the images from the historical military aerial photographing taken in 1952, which depicts the form of the village Kapličky before its complete destruction (Figure 5). The macroscopic view of the landscape around Kapličky (Kramer, 1980) allowed to identify the relics of the village, isolated farmsteads, roads, ponds and field systems. It is interesting that the aerial photographs show a linear course of the sand zone, which was part of the first construction phase of the Iron Curtain. The DTM does not show the sand stripe, but clearly draws the course of the terrain edge, which is the relict of the wire signal wall barrier from the 1980s once equipped with the electric-wire fence.

Further concretization of village remains, such as individual buildings, followed. Correlation of the Digital Terrain Model and the aerial image with the detailed analysis of cartographic sources (military mapping (Figure 6), cadastral maps (Figure 7) and plans) and field survey of the site. As a result, we could identify individual remains, and link structures and sites with their names, the names of the original owners, crafts they were working in (shoemaker, carpenter, grave-digger, etc.), and communal institutions (school, etc.) (Figure 8).
Figure 6: Aerial photography of Dobřín and Kapličky from 1952.

Source: Historical orthophoto. © CENIA 2010 and GEODIS BRNO, spol. s r.o. 2010. Underlying aerial photographs were provided by VÚHMÚ Dobruška, © MO ČR 2009

Figure 7: Scheme of Kapličky village (Kappeln) on a stable cadastre from the 1830s.


Figure 8: Identification of individual buildings from Kapličky on photograph of 1930.

Source: visualization by M. Preusz, data collection J. Doktorová, M. Zvánovec

Figure 8: Identification of individual buildings from Kapličky on photograph of 1930.
The view of Kapličky from the bird's-eye perspective can be supplemented by the information drawn from historical photographs. Geo-referencing of historical photos allows us not only to establish the point, where the photographer stood, but also to compare the past and present of human activities in the landscape by juxtaposing the historical image with contemporary photograph (Figure 9).

In the following step, 3D model reconstruction of the destroyed village buildings or isolated farms and structures was created with the help of ground plans obtained from the archaeological prospection, the historical plans, aerial and LiDAR photographs and historical photographs. Among the buildings, St. Peter and Paul church in Kapličky, which was the dominant of the village and surrounding countryside, has been virtually reconstructed (Figure 10). The Czechoslovak army blew up the building in 1959, but thanks to the witness reports, historical photos and descriptions, we possess valuable information about its exterior and inner equipment.

Results and discussion

As shown in the case study, the main aim of the project is the long-term archiving and public accessibility of historical photographs that can be used in the research and reconstruction of destroyed historical and archaeological sites, extinct settlements and landscape transformation. Reconstruction of Kapličky village and its surroundings not only proves the benefits of using historical photographs in research and reconstruction of lost sites, but also the advantages of joining computer-based technologies when applied to photo processing. The result is a multi-step process of automatized image recognition, research data processing and public response.

On the organisation level, the project is based on interdisciplinary collaboration of IT experts with historians, archaeologists, and natural scientists, while a broad public can cooperate in the process of analysis and location
Geoinformatics and Crowdsourcing in Cultural Heritage: A Tool for Managing Historical Archives

of historical photographs on the map through tagging and crowdsourcing. The project is now running in its second year, but it has already contributed to a better understanding of the transformation of cultural and natural landscape elements in the Šumava landscape. As an obvious asset, the results of the work are directly accessible to the broad public via web portal (users can immediately browse both the research and crowdsourced data, distinguished from each other by specific marking) and will be better understood by it, as it directly participates in their collection.

On the research outcomes level, long-term archived and geo-referenced photographs have undeniable importance for in preserving local memory and cultural or natural heritage, especially in the former conflict zones. In the Czech border region this interest in lost sites and settlements started to evolve, after 1989 (Bureš, 2015) due to political changes after the fall of Iron Curtain. Using historical photographs as source of information with the help of automatized and IT tools enables to handle and mine large pool of data and diverse material (both textual and visual) to understand and reconstruct landscapes, which have undergone turbulent changes, sometimes due to tragic historical events, during the past one and half century. The suggested IT-based methods help to re-install historical photographs to their role as valuable witnesses of human past.

Conclusion

Alone standing historical photographs are valuable witnesses of human past. But enriched by detailed information about their content and even about the circumstances of the acquisition can tell us important stories. This article aimed to present current state of art of ongoing project PhotoStruk. The scope of this project is to develop tools and methodology for work with historical photographs hidden in archives. But also to indicate possibilities of their use for different science.

Acknowledgments

This article was written with the support of the project "PhotoStruk - Analysis of Historical Photographs for Virtual Reconstruction of the Cultural Heritage in the Czech-Bavarian Border", No. 63 within the Cross-Border Cooperation Program Freistaat Bavaria - Czech Republic Objective ETZ 2014-2020.

Corresponding authors:
Dott.ssa Klára Paclíková
Institute of Applied Informatics, Faculty of Science, University of South Bohemia Ceske Budejovice
Braníkovská 1760, 370 05 České Budějovice, Czech Republic
Phone: 00420 773 483 012, E-mail: klara.paclikova@gmail.com

References


