https://doi.org/10.2352/issn.2168-3204.2020.1.0.79

Artificial Intelligence for content and context metadata retrieval in photographs and image groups.

Peter Fornaro, Vera Chiquet; University of Basel, Digital Humanities Lab; Basel, Switzerland

Abstract

Digitization projects of analog photographic collections are still growing in number, and therefore such assets of images become bigger continuously. Also, there is a strong trend towards open data and interfaces to access and reuse the image resources (FAIR data). To be able to search and find images in a repository, metadata of a certain depth must be existing. Typically, indexing and valorization, done by experts that know the (photographic) collections, is necessary to achieve such meta-information. There are various metadata standards based on different concepts for the description of collections. Some, like ISAD(G), are more related to the physical structure of archives, others, like CIDOC-CRM, take into account the content of the images in detail. Enhancing the depth of indexing increases the time necessary drastically. It is also a task that is not easily scalable because specific content related knowledge is necessary. With the assistance of artificial intelligence, historic photographic collections could potentially be enhanced with metadata semi-automatically. For the successful application of machine learning, it is essential to have robust training sets.

In the presented paper, we show our observations in monitoring participants indexing historic collections of photographs. In the observations of workshops of people working with photographic heritage, it was monitored how single photographs but also image groups are described. Based on that knowledge, machine learning components can be trained and optimized for that particular type of source material. The demonstrated approach has the potential to support the work of valorization substantially. In addition, the approach has, to some extent, the potential to preserve the fundamental structures of knowledge of contemporary witnesses.

Motivation

In the interdisciplinary field of Digital Humanities, the symbiotic interaction of image and text is of particular interest. In the context of photographic collections, descriptive metadata is, for example, important to find and identify photographs and set them e.g. in relation to each other. Projects to scan photographic images are still popular and they are growing in number. Optimized workflows and robust capture hardware allow high imaging throughput, without a reduction in quality. Historic photographs have a limited lifetime. A solution for preservation is digitization whereby hundreds of images can be scanned per person and day. As a consequence, the number of digital images generated and stored in digital repositories is growing fast. For finding, accessing, and the reuse of such items of photographic collections is metadata essential.

State-of-the-art digital infrastructures do not only guarantee the preservation of the digital files for longterm accessibility, but they also support complex metadata schemas, offer a framework for ontologies and allow to search for combinations of keywords and semantic relations [1]. Open access becomes more and more crucial. Standardized interfaces (Application Programming Interfaces, APIs) allow in a second step the controlled access to the resources for interoperability with other repositories and e.g. graphical user interfaces for virtual interaction. However, the presence of actual metadata, text-based describing content, is essential for the functional application of such a digital database environment. Extensive collections of photographs without appropriate metadata are not functional, not interoperable, finally, it doesn't fulfill the requirements of FAIR data [2]. Initiatives like the International Image Interoperability Framework (IIIF) [3] show how vital functional data exchange is [4]. Photographic collections without metadata are, to some extent, just data bunkered in digital silos. Metadata is essential to make image collections functional but metadata needs to be generated or acquired. The presented approach helps to make use of photographic collections in the digital domain. It is, to some extent, fundamental for a generic and scalable solution to catalog photographic collections.

Problem

Photographs are important sources for many fields of research. In art history photographs allow for example easy dissemination of the visual appearance of paintings or sculptures. In archeology, photographs are often the only way to transport information from the excavation back to the home institution and in social science photographic images allow e.g. to trace historic events. The rich visual information of photographic images is the basis for the use of such images as scientific and historical sources. However, it is not easy to acquire the information they contain in order to make such photographic images reusable for new research. Without this essential meta-information, most images are probably aesthetic and appealing visual representations but lack defined facts such as date, occasion, intention or just a description of the scene represented. The original purpose and context of photographs are, in most cases, not directly given due to the lack of written information on the photographs itself or additional documents passed down. Necessary metadata such as where and when an image was taken is essential in order to be able to use photography as a scientifically relevant source. In addition to understanding the image content and context, which may be

elaborated out of the relations in image groups, is essential for many fields of research in humanities and social sciences. There is another issue arising: Photographic images are often composed in albums or in collections. The necessary metadata to capture those aspects of "belonging together" or "set in context" is not easy to capture but probably it is an essential further aspect for research with and about photographs. For example in the case of very large assets of images, a subset that stays in relation can help to structure and prioritize the process of cataloging. Additionally, the dissemination and presentation of prepared photographic stories is probably an approach to attract people to e.g. enjoy the precious images of our history. Even if open access has grown in the last years, there is still the potential of invisible data silos due to the lack of stories told. The Europeana platform is a good example of a successful initiative to promote photographic material of all kinds.

However, the enrichment of digital photographs with metadata about the scene, the provenience and the context is up to today not as performant and scalable as the process of image capture. Usually, the first stage of digitization and the later cataloging is separated completely in regards to the expertise of the human resources and the time of the event. Especially finding and defining meta information can be very time-consuming. On the other hand, there is a rich variety of metadata standards that can be used, ranging from the rather less complex DC Dublin Core, over archive specific standards like ISAD(G), or the comprehensive CIDOC-CRM. In addition, we have experienced in various cases that such standards like e.g. the ones mentioned above are modified to be used as extended house schemas. Besides such structures for metadata, existing vocabularies to ensure the correctness of wording and labeling of e.g .places, objects or persons. It is clear that such a process of cataloging is time-consuming and intense in regard to the required know-how.

There are a number of different approaches for speeding up the process of indexing of photographic images and image collections. The simplest approach is to reduce the complexity of the metadata. Non-image-specific metadata can be assigned automatically. Typical information of this kind are a copyright notice or the name of a photographer who was responsible for an entire collection. The reduction of metadata complexity might result in a reduction of the quality of the source material. Especially in the digital domain, the consideration of metadata should be part of source criticism. Another approach, that became popular more recently as a side effect of interoperable data, is crowdsourcing. Crowdsourcing tries to collect information from a large number of persons. In the case of crowdsourcing, quality control is certainly an issue. To ensure the quality of such meta-information of people with unknown expertise concepts like *Output Agreement, Contributor Evaluation* or *Expert Review* can be applied [11]. Especially for historic photographic material, crowdsourcing is very promising, because of the depth and broadness of the know-how that becomes accessible by the Web [12].

Last but not least computer-aided solutions like image-based content retrieval are very interesting for the generation of metadata. Especially noteworthy is the capability of artificial intelligence to support the cataloging of single images but also of image collections to gain more complex metadata. Such computer-generated meta-information about photographic image collections is a chance to work more efficiently with large assets of photographic images, that are not fully understood and valorized. Especially the reproducibility of results of computer-aided approaches is an advantage of such a solution. The appearance of systematically wrong metadata can be used to improve the results by enhanced training data. There is another aspect that is interesting: Especially in the field of historic photography contemporary witnesses usually know the picture contents in more detail and more correctly. From a Digital Humanities perspective, it is an opportunity, if not an obligation, to explore and evaluate the potential for recording and digitally preserving the know-how on historic photographs carried out by the knowledgeable generation.

Metadata about photographs can be gained from various sources: Labelings and stamps potentially give information about photographers, studios, persons in the scene or cities captured. But even if metadata is given as text, for the transformation into the digital domain, software components like optical character recognition (OCR) are necessary to transform the image of a written meta information into digital code, digital text that can be worked with. Without such explicit written information, the content of the image must be analyzed, to define the objects or sites captured. Content-Based Image Retrieval (CBIR) can mean various approaches to recognized image patterns or specific structures that can be tagged and described with digital meta text [13]. Today CBIR can be found in many devices, like smartphones, to automatically catalog images stored in applications, like Photos running on macOS. Words like "Trees", "House", or even "Birthday" can be searched for without ever having tagged the images manually. On the other hand are the capabilities to recognize complete scenes rather limited, especially if objects shown look different on historic photographs than today. A simple test shows this effect of e.g. different designs. If an image collection of 40'000 random pictures, stored in Apple Photos, is searched for "Telephone", only vintage stereotype looking apparatuses are found but not any modern device for communication like mobile phone or smartphone. This limitation to already known knowledge, in other words, training sets, gets even more drastic if the metadata becomes complex, like in the case of semantic description. Such a degree of complexity can be found in related image clusters, as found in albums. Albums of photographs are actually meant to tell stories, composed of more than just pure image content; they are collections of images with a narrative, a set of meaningfully related images that belong physically and in many cases contextually together. For any CBIR system, such a related set of photographs is a challenge. For the successful test of such an approach reference sets of well-known and rich described photographs are essential. To get such training data monitoring of peer groups, while describing such photographic content is an approach we have followed as a first experiment.

The digital representation of such results can be done with semantic approaches [5]. They allow representing complex structures of meta-information like found in albums, leading to meaningful descriptions and the potential to make e.g. use of semantic reasoning to find correlations that are not obvious in the first instance. To ensure quality metadata, it is essential to evaluate all of the approaches. It needs various but different technologies (CBIR, adequate vocabularies, metadata standards that allow the representation of semantic relations) to achieve automatically generated information of good quality for the use in research. Especially the description of related image collections, like photographic albums, is still a challenge in cataloging. There is also a lack of easy to use solutions, embedded in existing digital database infrastructures to make use of sophisticated solutions for semi-automated image-based metadata extraction.

Intermediate Results

In the framework of a European project in conjunction with Photoconsortium, a Europen network of Experts in Photography, a workshop about the indexing and valorization of photographic assets was held. In the workshop, two different peer groups were brought together; representatives of the grandparent generation and the grandchild generation that were not related.

Participants were encouraged to bring their family photographs. The main interest in the workshop was to focus on the description, the indexing work and the generation of metadata about the photographs brought along. The results of such an experiment give insights into the methodology while speaking about single pictures or image clusters, digital or historical photographs.

CBIR, which is used for keyword searches for images, can be helpful to use images as scientific and historical sources with metadata. Today's CBIR may work well for current photographs, but it is more difficult for historical photographs. Here it needs to be developed and examined further for the indexing of historical, and especially family photography. This workshop showed that the younger generation without witnessing knowledge about the story behind the photography tried to identify symbols in the pictures or describe poses, clothing, architecture, backgrounds and any signs that could help identify time or place. They were also able to propose social aspects and cultural conventions and to point out meanings that the witnesses are generally not aware of. By contrast, only the witnesses can contextualize the historical photograph. So it is very fruitful to combine the position of the knowing witness with the distant observer for retrieving metadata. Because these historical photographs support not only the living memories of the individual but also help to re-make sense of the past as well as the present, the circumstance and the network they were produced but also looked at and discussed: "Formally as well functionally photographic memory pictures belong to both spheres: the individual as well as the collective, the private as well as the public" (Ruchatz 2018: 200) [6].

If we do not gather this information about the photographs of this generation and do not ask about the meaning of historical photographs, sooner or later the background and the history will be lost and will no longer be retrievable like this. With the witness generation, we can access from the communication exchange more metadata concerning social aspects, context and meaning. Browsing a photo album or describing image clusters with both groups can generate appropriate metadata for machine learning. To train this methodology, we use the knowledge of the generation that knows in detail what is shown in the old photographs. The vintage photographs and the appropriate metadata define our reference set to train artificial solutions. The content can also only be grasped with the people who can still recognize the things in the old photographs and describe their meaning and relation or function to other photographs.

It was interesting to observe that the description of photo groups, the comparison, linking, and extended viewing, as well as the discussion of several grouped photographs, represents a high added value. Describing groups of images can be very interesting, especially for a better understanding of contextual metadata. We want to use the information from the peers not only for the specific crowd science on the exploitation of those unique collections, but to re-use the gained information also for other collections by the application of the trained AI approach. The peer input is used for the training of AI for the specific case of historical family photographs and image clusters.

Approach

For photographic sources of high value, we want to use simple metadata, but also more sophisticated descriptions, which relate single images with other images, make them comparable, or affiliate them to a cluster of other related photographs. Especially the approach of image clusters is a promising approach for machine learning. Conventional CBIR done on single images can be regarded as a close-up evaluation of photographic assets, whereas the analysis of clusters of photographs can be regarded as a macroscopic perspective. If an AI system is trained by people who were witnesses of time or have the knowledge of historical context, it becomes possible to store such cognitive and semantic aspects in a digital system – a fundamental step towards a new level of information preservation. In the interdisciplinary project, we go beyond the commonly used image retrieval functionality. We want to use machine learning to gain contexts and the semantic networks between the photographs.

The artificial intelligence components will be applied and evaluated as follows:

1. In the project, the state-of-the-art approaches for CBIR will be used to apply them to historic photographic collections. Such collections show specific attributes like healthy photographic grain, material defects, faded colors, or irregularities in tonality. Based on well-chosen photographic samples, a CBIR engine will be examined and optimized for the assets of vintage images. The results will be analyzed, and the CBIR system will be optimized for this specific application [7].

2. Technologically modified CBIR approaches will be applied to image clusters. A cluster is a set of related photographs that show similarities or have a contextual relation. The cluster size can vary from 2 to n images. Images within a cluster can have various distances to each other, whereas the distance represents the strength of the relation. Multiple clusters have relations between them, as well. The attribute of distance is applicable to clusters-relations also. In this experimental approach, the feature set of multiple images will be defined, so that algorithmic processing is possible. The various attributes of image clusters will be explored and rated [8].

3. The meta-information gained in the workshops will be used to define an appropriate ontology. The ontology will be based on the existing work done e.g., by Getty, Matterhorn METS, or Cidoc-CRM. The ontology will be able to address the specific needs of the description of historical photographic material (close description, image facts) as well as of image context (distant description, cluster, and relation). The ontology must be able to represent all relevant information of a single image as well as image clusters. The increased dimensionality must be addressed with a dynamic, yet focused ontology [9].

The transfer from the machine learning output to the ontology will be examined for quality control of metadata. Only if the metadata output of an artificial component is of a minimum level of quality, the application in extensive collections makes sense [10].

Conclusion

The digital environment invites us to explore the capabilities of digital technology for data structuring, segmentation, and to improve accessibility and usability. Especially photographic images are an essential part of our cultural heritage. To understand the actual practices, we have to look back at historic photographs. What changed and what stayed the same, before and after the transition from analog to the digital domain. We observe and analyze the work with analog photographic collections such as the interaction between humans and images as well as the potential to employ digital technologies to improve the methodology in work with images and to transfer it to artificial intelligence systems.

The application of machine learning to historic assets is polarizing. From humanities researchers, the application of computer technology to historic assets can only hardly succeed, due to the lack of in-depth knowledge of the complicated matter of photographic collections.
With this project, we combine the exploitation of humanistic cataloging (generation of metadata with witnesses) with AI to protect an essential piece of cultural heritage, making sure that artificial intelligence takes the role of support for the preservation of pre-digital knowledge. It is also an important project for the field of digital humanities to have a case for the dialog between interdisciplinary researchers – computer scientists and humanities scholars.

The presented approach can be a chance to preserve the knowledge of contemporary witnesses for future work with not yet digitized material; AI can be a solution here, at least to some extent. Therefore, the capabilities of AI for historic photographs and their relations must be explored and evaluated.

References

[1] Lukas Rosenthaler, Peter Fornaro, Claire Clivaz (2015). DASCH: Data and service center for the humanities http://dx.doi.org/10.1093/llc/fqv051 i43-i49 First published online: 8 October 2015.

[2] Lukas Rosenthaler, Andrea Bianco, Peter Fornaro (2017), Integrating Image Resources Into Virtual Research Environments For The Humanities – a Simple Image Presentation Interface (SIPI) based on IIIF, Digital Humanities 2017 Conference, University of Montreal, Canada.

[3] Snydman, S.(2015) The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images, IS&T Archiving Conference Proceedings, Washington DC.

[4] Tobias Schweizer, Peter Fornaro, Lukas Rosenthaler (2017), Content-based Interoperability: Beyond the Merely Technical Specifications of Interfaces, IS&T Archiving Conference Proceedings, Riga.

[5] Sicilia, M.-A. (2006). Metadata, semantics, and ontology: providing meaning to information resources, Int. J. Metadata, Semantics and Ontologies, Vol. 1, No. 1, 2006.

[6] Ruchatz, Jens (2018). Public Rites/Private Memories. Reconciling the Social and Individual in Wedding Photography. In: Global Photographies Helff, S. and Micheli, S. (Eds.) transcript, Bielefeld, p. 177-202.

[7] Iakovidis, D. et al (2007). Pattern-Based Retrieval of Cultural Heritage Images. Proc. of the 11th Panhellenic Conference in Informatics (PCI'07).

[8] Luca Rossetto, Ralph Gasser, Silvan Heller, Mahnaz Amiri Parian, Heiko Schuldt (2019). Retrieval of Structured and Unstructured Data with vitrivr, Proceedings of the International Conference on Multimedia Retrieval (ICMR), Ottawa, ON, Canada 2019/6.

[9] Sandhaus, P., Boll, S. (2011). Semantic analysis and retrieval in personal and social photo collections In Multimed Tools Appl (2011) 51: 5. https://doi.org/10.1007/s11042-010-0673-1.

[10] Stork, D.G. (2000). Open data collection for training intelligent software in the Open Mind Initiative.

[11] Allahbakhsh, Mohammad & Benatallah, Boualem & Ignjatović, Aleksandar & Motahari Nezhad, Hamid R. & Bertino, Elisa & Dustdar, Schahram. (2013). Quality Control in Crowdsourcing Systems: Issues and Directions. Internet Computing, IEEE. 17. 76-81. 10.1109/MIC.2013.20.

[12] Graf, Nicole. (2014). Crowdsourcing – neue Möglichkeiten und Grenzen für Bildarchive. o-bib. Das offene Bibliotheksjournal. 1. 10.5282/o-bib/2014H1S249-253.

[13] Alexander, Malcom & Gunasekaran, S. (2014). A Survey on Image Retrieval Methods. 10.13140/2.1.2779.0723.

Author Biography

Peter Fornaro is member of the management team of the Digital Humanities Lab of the University of Basel. He is doing research in the field of digital archiving, imaging, cultural heritage preservation and computational photography. Fornaro is teaching at the University of Basel. Besides research and lecturing, he is giving consulting to companies, archives, and museums. Fornaro is a member of the Swiss Commission for Cultural Heritage Preservation (EKKGS).

Vera Chiquet works as a scientific assistant at the Digital Humanities Lab at the University of Basel and the Data and Service Center for the Humanities (DaSCH). She taught and researched in Basel, Hamburg, and Bologna. Her research focuses on mass media, photography, and digitization. She is president of the Friends of the Herzog Foundation, Laboratory of Photography.