Visual Analytics and Computer Vision Meet Cultural Heritage

Dissertation Topics

doc.funds.connect

Wolfgang Aigner (Coordinator), Franziska Bruckner, Markus Seidl, Matthias Zeppelzauer
Institute of Creative\Media/Technologies, St. Pölten University of Applied Sciences
Mathias Corvinus-Strasse 15, A-3100 St. Poelten, Austria
https://icmt.fhstp.ac.at

Silvia Miksch (Second Coordinator), Robert Sablatnig, Manuela Waldner
Institute of Visual Computing and Human-Centered Technology, TU Wien
Favoritenstrasse 9/193, A-1040 Vienna, Austria
https://informatics.tuwien.ac.at/orgs/e193

June 7, 2023

Dissertation Topic 1: Content / Time Prediction: Computational Analysis of Historical Ephemeral Films

Automatic analysis of contemporary time-based visual media is a long standing research field that fosters research activities ranging from low-level analysis tasks (such as shot boundary detection or near-duplicate detection) to complex spatio-temporal scene analysis tasks on object or even concept level (e.g., (Liang et al., 2020)). These activities have been evaluated in well-known benchmarks such as the TRECVid campaign, which has been held yearly since 2003[1]. In contrast to that, automatic analysis of historical film material has drawn way less attention. Additionally, it has been shown that methods and approaches dealing with contemporary video material cannot be transferred directly to historical material (Helm and Kampel, 2019; Seidl et al., 2011). In this dissertation, the PhD candidate will work on especially challenging historical material stemming from different sources, consisting of 80 ephemeral films from a time span of more than 90 years.

Aims Efficient scene analysis methods are a necessary pre-requisite for film historians to, on the one hand, allow quantitative analyses and, on the other hand, search for specific scene content or abstract scene relations within large film collections. The final goal of computational historical film analysis is to develop generally applicable approaches for spatio-temporal scene analysis and similarity estimation. The PhD candidate has the freedom to select any unsolved analysis task(s) to investigate in course of the dissertation.

Methods The spatio-temporal scene analysis approaches for historical material need to be robust against certain properties of the historical material, such as flicker, image vibrations, degraded contrasts, scratches, fungus, dirt, wrong exposure, or film development. Deep neural networks, e.g., with multi-stream architectures allowing for the detection of spatio-temporal relations (Feichtenhofer et al., 2016) in high-quality contemporary video material can serve as starting points for own investigations. Additionally, for low-level tasks, classical machine learning approaches might be an option. The material which we are using stems from the European H2020 project I-Media-Cities[2] and has been annotated on shot level. Hence, the doctoral candidate will not have to spend any additional efforts to prepare the material for the computational analysis tasks.

The Ludwig Boltzmann Institute Digital History (LBIDH) is acting as partner. On the one hand, it is providing the annotated film material. On the other hand, the director of the LBIDH, Dr. Ingo Zechner agreed to be a member of the scientific advisory board of VaCoViCu.

Time frame This dissertation project is planned to be finished within the funding period of the doctoral college (Year 1 – Year 4).

Faculty members Markus Seidl (FHSTP), Robert Sablatnig (TU Wien), and Franziska Bruckner (FHSTP) + Alexandra Schneider (associate faculty).

[1] https://trecvid.nist.gov/ (accessed June 7, 2023)
[2] https://www.imediacities.eu/ (accessed June 7, 2023)

Dissertation Topic 2: Intent Analysis: Determination of Camera Type, Lens Type and Camera Position for Historical Images

Albeit being generally attributed with the qualities of evidence for what in fact has happened, a photograph merely represents an extract or excerpt of reality, regardless of being putative or constructed. Exploring and analyzing such intents behind certain iconic photos is of great importance to the work of historians. Part of these intents can be traced back to certain attributes of used equipment (e.g., lens and camera type), image composition (e.g., shot type), processing (e.g., retouching and clipping), as well as camera position and orientation that can be analyzed using computational methods.

A challenging aspect is that historical photos are very different to modern photos. On the one hand, there are a number of approaches for digital images, however, these work only partially for historical images since, in general, the quality of the images is much lower than with modern images. On the other hand, since the number of cameras and lenses was very limited, photo historians are able to identify the camera and lens type of a photo through given contextual information of the scene such as place, time and historical event (like the burial of emperor Franz Josef).

Aims Given a large historical photo collection, the research question is if it is possible to do this classification automatically if context information is available and, in a second step, if this is possible even without context information as it is the case for digital cameras. Furthermore, the camera position relative to the scene should be also determined automatically to reason about the intention of the photographer while taking the photo.

Methods For shot type classification, modern deep learning-based classification architectures will be applied (e.g., VGG, ResNet or GoogleNet). Single-shot estimation of camera position and orientation will be done by applying projective geometry observations or Convolutional Neural Networks (CNN) in combination with publicly available datasets. For historical camera image forensics, a Deep Learning based architecture is aimed for. Of particular importance is a human-in-the-loop analytics approach that allows photo historians to annotate training data sets with known attributes in order to optimize the identification system (supervised learning). For this purpose, knowledge-assisted visual analytics methods will be applied (Federico et al., 2017).

For this topic, we will collaborate with the historical photo archive of the workers' newspaper (Austrian Labour History Society VGA) that comprises a large corpus of 600,000 historical images (see collaboration letters for the collaboration agreement with VGA).

Time frame This dissertation project is planned to be finished within the funding period of the doctoral college (Year 1 – Year 4).

Faculty members Robert Sablatnig (TU Wien) and Wolfgang Aigner (FHSTP) + Yvonne Zimmermann (associate faculty).

Dissertation Topic 3: Context Analysis: Exploratory Analysis of Context in Historical Images

Historical visual material implicitly contains rich information about the temporal development of objects, people, concepts and topics. For example, historians may analyze the role of traditional costumes in Austria over the last century, and non-expert "rail fans" may want to investigate how the public transport was co-notated over time. Through exploratory analysis, users may want to analyze the *context* in which the objects of interest have been depicted in historical photographs, as well as their relative frequency and how the context has changed over time. Such user requests are, however, highly subjective and thus no readily available retrieval models exist. Furthermore, there is usually a lack of suitable metadata in historical collections for such specific and sophisticated tasks (Windhager et al., 2018).

Aims The goal of this PhD topic is to design, develop, and evaluate a novel community-driven interactive visual analytics approach for the exploration and analysis of *context* in sparsely labeled image collections. The core component will be an interactive visual interface that supports the exploration of large media collections and helps the users to discover concepts of interest. A multi-label image classification model shall be constructed that incrementally learns user-specific concepts of interest and thereby relevant metadata categories. The (online) trained model, in turn, shall support the users' exploratory analysis by useful personalized recommendations matching the users search intent (Kofler et al., 2016) and visually summarizing the predicted context of the retrieved images over time, while taking prediction uncertainties into account.

Methods The novel approach for analyzing semantic visual context changes over time shall build upon and extend visual interactive learning (Bernard et al., 2018) as well as methodology from linguistic research, where diachronic word embeddings are analyzed to reveal semantic changes of words over time (Hamilton et al., 2016). The candidate will develop novel methods to apply this principle to large, time-dependent image collections with sparse labels. The student will also investigate methods to visualize the analysis results with a special focus on the inherent uncertainty given the sparse and noisy tags assigned to historical images. The major contribution will be a novel human-in-the-loop method for image-based interactive context analysis and concept discovery, where the machine will incrementally learn user-specific tags and recommend potentially interesting content. Evaluation will be a mixture of quantitative performance evaluation with respect to ground truth as well as qualitative user studies to assess the usefulness of the new proposed technique.

We will work with data provided by ICARUS via the Topothek Austria[3] and the Photo Institute Bonartes[4] (letters of support attached in Annex **??**).

Time frame This dissertation project is planned to be finished within the funding period of the doctoral college (Year 1 – Year 4).

Faculty members Matthias Zeppelzauer (FHSTP) and Manuela Waldner (TU Wien) + Yvonne Zimmermann (associate faculty).

[3] https://www.icar-us.eu/en/cooperation/online-portals/topothek (accessed June 7, 2023)
[4] https://www.bonartes.org, (accessed June 7, 2023)

Dissertation Topic 4: Relationship Analysis: Visual Exploration of Networks in Time and Space

Cultural objects (e.g., photographs, film) capture relationships over time and space, which can be seen as dynamic networks. Relationships are described through metadata, such as locations, events, time, or similarity relations. However, these relations contain severe data quality problems and uncertainty. Content-based information retrieval methodology (Zhou et al., 2017) will be utilized to detect and curate appropriate relationships, which will then be integrated into an interactive visual exploration environment for complex networks to support domain experts (historians) during their task-specific analysis process (Raieli, 2016).

Aims The goal of this thesis topic is to explore relationships of cultural objects by designing and evaluating visual analytics methods to tackle such networks in time and space and to support the domain experts in the field of digital art history to gain insights into changes over time and space. The overall vision is to intertwine interactive visualization, content-based information retrieval, and recommendations with a particular focus on complex networks representing relations in time and space. A further challenge is to scale such a visualization and exploration approach to large-scale corpora of thousands of images and, at the same time, making the visualized data easy to grasp by the user. This PhD topic complements topic T3 (see above), which focuses on *context*. Here, the focus is on *objects* and their relations over time and space.

Methods From a methodological perspective, we follow a human-centered and data-driven research process based on the design triangle (Data–Users–Tasks paradigm) by (Miksch and Aigner, 2014) enriched by the design study methodology by (Sedlmair et al., 2012) and the nested model for visualization design and validation by (Munzner, 2009). Node-link diagrams are widely used approaches to visualize, interactively explore, and analyze dynamic networks. Other approaches apply (adjacency) matrices, space-filling (i.e., maps), and hybrid approaches (Beck et al., 2017; Nobre et al., 2019). However, these approaches do not utilize the content of the cultural objects as well as the specific tasks of historians. According to the proposed data and tasks abstraction, we aim to provide multimodally enriched visual analytics solutions for domain experts, such as historians.

Similar to topic 5, we will work with data provided by the International Centre for Archival Research (ICARUS) via the Topothek Austria[5] as well as the Institute of Balkan and Habsburg Studies at the Austrian Academy of Sciences (ÖAW). In addition we will use the handcrafted dataset from the IMMV project[6]. Collaboration letters of both organizations are attached.

Time frame This dissertation project is planned to be finished within the funding period of the doctoral college (Year 1 – Year 4).

Faculty members Silvia Miksch (TU Wien) and Matthias Zeppelzauer (FHSTP) + Arno Strohmayer (associate faculty).

[5] https://www.icar-us.eu/en/cooperation/online-portals/topothek/ (accessed June 7, 2023)
[6] http://www.musicmapping.at (accessed June 7, 2023)

Dissertation Topic 5: Situated Storytelling: History in your Pocket – Visual Storytelling in Space and Time

Biographical records and event timelines are invaluable and important sources for historical research. However, their accessibility and comparability is limited. Exhibits and artifacts in museums and archives are often related to certain locations in the real world (e.g., a photo of a specific city, the hometown of an artist, an archaeological object that was found at a certain location, etc.). Even though such information is provided in a museum, virtual museum or archive, it is hardly ever made visible at the actual geospatial position outside the museum.

Aims The aim of this thesis is to design, implement, and evaluate methods that relate digital information to physical spaces in form of situated mobile visualizations (Thomas et al., 2018). Moreover, the dominate linear narrative modes of history is limited according to various exploration perspectives. This subproject aims to design an interactive visual exploration environment to access and understand fundamental influences of sociopolitical and economic factors in history and to enrich visualizations with associative information with respect to time and space.

Methods We will follow a human-centered and data-driven research process based on the nested model for visualization design (Munzner, 2009) which closely intertwines design and evaluation phases while mitigating potential risks and threats for validity. It is a unified approach that splits visualization design into four levels and combines these with appropriate evaluation methods to mitigate threats to validity at each level. Starting from the top, the levels are domain problem and data characterization (understanding problem domain and users' tasks and goals), operation and data type abstraction (transform domain-specific description and raw data into a more generic notion; in this case biographical/timeline data with spatial anchoring and multivariate multimedia data), visual encoding and interaction design (design of appropriate methods to represent and interact with data; in our case we mainly focus on methods for see-through AR devices such as smartphones and tablets), and algorithm design (design of specific implementation of methods, where we will utilize web technologies to ensure broad accessibility). To account for the specifics of situated visualization, we will follow the conceptual models of (Thomas et al., 2018) and (Willett et al., 2017) that describe the interplay of logical and physical world. Localization of device positions will be achieved using image or object markers (mainly for indoor situations) or position and orientation information provided by sensors of the mobile device. The data we will be working with are historical photographs as well as structured event data. Specifically, we will work with data provided by the International Centre for Archival Research (ICARUS) via the Topothek Austria[7] as well as the Institute of Balkan and Habsburg Studies at the Austrian Academy of Sciences (ÖAW), whose historians digitally edit several important historical collections. These include image representations of sources and contain curated person, institution, and place entity data (see collaboration letters for further details).

Time frame This dissertation project is planned to be finished within the funding period of the doctoral college (Year 1 – Year 4).

Faculty members Wolfgang Aigner (FHSTP) and Silvia Miksch (TU Wien) + Susana Zapke (associate faculty).

[7] https://www.icar-us.eu/en/cooperation/online-portals/topothek/ (accessed June 7, 2023)

References

Beck, F. et al. (2017). "A taxonomy and survey of dynamic graph visualization". In: *Comp. Graph. Forum* 36.1, pp. 133–159. DOI: 10/f9wkd9.

Bernard, J. et al. (2018). "VIAL – a unified process for visual-interactive labeling". In: *The Visual Comp.* 34, pp. 1189–1207. DOI: 10/gd5hr3.

Federico, P. et al. (2017). "The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics". In: *Proc. VAST*. IEEE. DOI: 10/ghppzr.

Feichtenhofer, C., A. Pinz, and A. Zisserman (2016). "Convolutional two-stream network fusion for video action recognition". In: *2016 IEEE CVPR*, pp. 1933–1941. DOI: 10/gf2j4h.

Hamilton, W. L., J. Leskovec, and D. Jurafsky (2016). "Diachronic word embeddings reveal statistical laws of semantic change". In: *Proc. meeting of the assoc. for comp. ling.* Assoc. for Comp. Ling., pp. 1489–1501. DOI: 10/gfw6bx.

Helm, D. and M. Kampel (2019). "Shot boundary detection for automatic video analysis of historical films". In: *New trends in image analysis and processing – ICIAP 2019*. Springer, pp. 137–147.

Kofler, C., M. Larson, and A. Hanjalic (2016). "User intent in multimedia search: a survey of the state of the art and future challenges". In: *ACM Comp. Surveys* 49.2, pp. 1–37. DOI: 10/ghpp3f.

Liang, Z., Y. Guan, and J. Rojas (2020). "Visual-semantic graph attention network for human-object interaction detection". In: *arXiv:2001.02302*. DOI: arXiv:2001.02302v5.

Miksch, S. and W. Aigner (2014). "A matter of time: Applying a data-users-tasks design triangle to visual analytics of time-oriented data". In: *Comp. & Graphics, Sp. Section on Visual Analytics* 38, pp. 286–290. DOI: 10/f3szvk.

Munzner, T. (2009). "A nested model for visualization design and validation". In: *IEEE TVCG* 15.6, pp. 921–928. DOI: 10/dzqbc2.

Nobre, C. et al. (2019). "The state of the art in visualizing multivariate networks". In: *Computer Graphics Forum* 38.3, pp. 807–832. DOI: 10/ghpp2g.

Raieli, R. (2016). "Introducing multimedia information retrieval to libraries". In: *JLIS. it* 7.3, pp. 9–42. DOI: 10/ghp4j9.

Sedlmair, M., M. Meyer, and T. Munzner (2012). "Design study methodology". In: *IEEE TVCG* 18.12, pp. 2431–2440. DOI: 10/f4fv7x.

Seidl, M. et al. (2011). "Gradual transition detection in historic film Material—a systematic study". In: *J. Comput. Cult. Herit.* 4.3. DOI: 10/fzsqr8.

Thomas, B. H. et al. (2018). "Situated Analytics". English. In: *Immersive Analytics*. LNCS. Springer, pp. 185–220. DOI: 10.1007/978-3-030-01388-2_7.

Willett, W., Y. Jansen, and P. Dragicevic (2017). "Embedded data representations". In: *IEEE TVCG* 23.1, pp. 461–470. DOI: 10/gfz688.

Windhager, F. et al. (2018). "Visualizing uncertainty in cultural heritage collections". In: *EuroVis workshop on reproducibility, verification, and validation in vis.* DOI: 10/ghp4j7.

Zhou, W., H. Li, and Q. Tian (2017). "Recent advance in content-based image retrieval: A literature survey". In: *arXiv*. DOI: arXiv:1706.06064.