

Multimodale KI zur Unterstützung geschichtswissenschaftlicher Quellenkritik Ein Forschungsauftritt

Muenster, Sander

sander.muenster@uni-jena.de
FSU Jena, Germany

Bruschke, Jonas

jonas.bruschke@uni-wuerzburg.de
JMU Wuerzburg, Germany

Kroeber, Cindy

cindy.kroeber@uni-jena.de
FSU Jena, Germany

Hoppe, Stephan

stephan.hoppe@kunstgeschichte.uni-muenchen.de
LMU Muenchen, Germany

Maiwald, Ferdinand

ferdinand.maiwald@uni-jena.de
FSU Jena, Germany

Niebling, Florian

florian.niebling@uni-wuerzburg.de
JMU Wuerzburg, Germany

Pattee, Aaron

aaron.pattee@lmu-muenchen.de
LMU Muenchen, Germany

Utescher, Ronja

ronja.utescher@uni-bielefeld.de
FSU Jena, Germany; U. Bielefeld, Germany

Zarriess, Sina

sina.zarriess@uni-bielefeld.de
U. Bielefeld, Germany

Einleitung

Fotografien und andere Abbilder von Architektur dienen in vielen historischen Wissenschaften als Quelle und Grundlage für fach- und theoriespezifische Untersuchungen. So werden zum Beispiel historische Fotoaufnahmen herangezogen, um den Zu-

stand eines Gebäudes zu rekonstruieren oder die Formensprache einer Epoche zu identifizieren. Ausgangspunkt dieser Szenarien aus Architektur-, Kunstgeschichte und Kulturwissenschaften ist eine durch Hilfsmittel der jeweiligen Fächer unterstützte Quellenrecherche und -kritik, auf die weitere Auswertungen und Verwendungen im wissenschaftlichen Kontext aufbauen.

Obwohl sich KI-basierte Methoden der *Computer Vision* in den letzten Jahren wesentlich weiterentwickelt haben, können diese den Prozess der Quellenrecherche und -kritik bisher allenfalls im Ansatz unterstützen, bspw. für die Exploration von Bildrepositorien oder das Retrieval von Bildern. Dies liegt zum einen daran, dass elementare diesbezügliche Vorgehensweisen zwar gut dokumentiert sind, WissenschaftlerInnen aber sehr individuell vorgehen. Zum anderen ist KI-Bildverarbeitung bisher wenig darauf ausgelegt, bildliche Inhalte multimodal zu kontextualisieren, d.h. verschiedene Quellengattungen wie Bilder und Texte zu kombinieren. Existierende Verfahren der Computer Vision extrahieren rein visuelle Merkmale und klassifizieren diese, während Texte oder Metadaten und darin enthaltenes Wissen wie bspw. Hinweise auf zeitliche Kontexte oder einzelne Motive nicht mit der Analyse verknüpft werden können.

Das BMBF-geförderte Projekt *HistKI* startete im Januar 2021 und will die Unterstützung und Modellierung von Bildquellenrecherche und -kritik als komplexe und grundlegende geschichtswissenschaftliche Arbeitstechnik durch multimodale KI-basierte Verfahren erforschen. Damit verbundene Teilfragen sind beispielsweise: Wie finden und beurteilen Historiker und andere Fachwissenschaftler Bildquellen? Welche generischen Vorgehensweisen und Teilproblemstellungen lassen sich hierfür identifizieren? Wie lässt sich dies mit KI-basierten Ansätzen befördern? Wie wirken sich KI-Techniken auf den geisteswissenschaftlichen Forschungsprozess aus?

Forschungsstand

Ausgangspunkt des Vorhabens ist der geschichtswissenschaftliche Forschungsprozess im Umgang mit Bildquellen. Für ein forschendes Handeln in den historischen Wissenschaften prägend sind einzelne Themenfelder (bspw. Kunst, Technik, Wirtschaft, Politik) sowie die darauf bezogenen Erkenntniszugänge. Leitendes Paradigma ist eine konstruktivistische Problemorientierung und damit eine quellenkritische, gegenstandsbezogene Analyse (Wengenroth, 1998, Reich, 2006). Methodologische Zugänge dazu liefern bspw. Hermeneutik, Semiotik (Holenstein, 1988, Zöllner, 2005) oder Phänomenologie (Pechtl, 2002). Grundsätzlich findet ein Zugang über Quellen statt und die damit verbundene Quellenkritik ist eine grundlegende Vorgehensweise historischer Forschung (Opgenoorth, 1997).

Modellierung forschenden Handelns: Eine Systematisierung von digitalem Forschungshandeln in den historischen Wissenschaften erfolgt bspw. aus Perspektive der eScience (Köhler et al., 2016) oder Informationswissenschaften, Ansätze zur Modellierung von Prozessabfolgen forschenden Handelns sind bspw. Forschungsprimitive (Münster and Terras, 2020, Kamposiori and Benardou, 2011). Demgegenüber stehen Ansätze zur sozialwissenschaftlichen Analyse des Informations- und Forschungsverhaltens (Münster et al., 2018, Ying et al., unpublished), bspw. der Science and Technology Studies (STS) u. a. anhand epistemischer Kulturen (Knorr-Cetina and Reichmann, 2015) oder der Wissenschaftsphilosophie anhand der Erkenntnisprozesse (Fleck, 1980, Popper, 1998). Damit gehen Ansätze zur Operationalisierung einerseits von der (1) „mechanistischen“ Idee aus, forschendes Handeln als Abfolge von operationalisierbaren Forschungsschrit-

ten (e.g. Forschungsprimitive, eScience) zu betrachten. Andererseits steht (2) die Vorstellung eines geisteswissenschaftlichen Forschungshandelns als „Black Box“ und erfahrungsbasiert, auf implizitem Wissen (Polanyi, 1966) basierend sowie der Wissensgenese als „serendipity“ (e.g. STS). Nicht zuletzt hinsichtlich der Lehrbarkeit haben sich in den bildbezogenen historischen Wissenschaften (3) semi-operationalisierte Zwischenformen etabliert – bspw. die bereits benannte Quellenkritik sowie speziell für Bildmedien die ikonologische Analyse (Panofsky, 1939). Trotz derartiger systematisierender Ansätze erfolgt eine Recherche und kritische Betrachtung von historischen Quellen in der Praxis hochgradig individuell und erfahrungsbasiert (Brieber et al., 2014, Münster et al., 2018). Vor diesem Hintergrund werden KI-basierte Ansätze derzeit vor allem zur Unterstützung von abgegrenzten Teilproblemen, wie Suche nach ähnlichen Bildern (Münster et al., 2019, Bell and Ommer, 2019), die Anreicherung von textuellen Metadaten (Lee and Münster, 2018) sowie Musteranalyse (Kohle, 2018, Klink, 2016), eingesetzt.

Bildwissenschaftliche Zugänge: Eine Reihe von aktuellen Publikationen und Initiativen greifen das Zusammenspiel von bildbezogenen historischen Wissenschaften und KI-basierten Forschungsansätzen auf – bspw. der Band „Digital Art History“ (Kuroczynski et al., 2019) sowie das DFG-Schwerpunktprogramm „Das digitale Bild“ (Kohle, 2018). Trotz heterogener Ansätze zur Bildrecherche und Quellenkritik existiert eine Reihe allgemeiner Problemstellungen (Hoppe and Breitling, 2016). HistKI liefert mit seiner inhaltsgetriebenen und forschungsprozessorientierten Fragestellung eine wichtige und komplementäre Grundlage für diese Initiative.

Language & Vision: Neuere Ansätze der Bildverarbeitung aus dem Bereich des Deep Learning ermöglichen nicht nur eine bessere Objekterkennung, sondern erweisen sich als besonders geeignet für *transfer learning* an der Schnittstelle von Bild- und Sprachverarbeitung. Dabei werden zum Beispiel semantische Repräsentationen wie *word* oder *sentence embeddings*, die aus Texten gelernt werden, anhand von multimodalen Daten wie Bildbeschreibungen mit visuellen Repräsentationen angereichert. Damit ist das Vokabular eines Objekterkennungssystems aus der Bildverarbeitung, das typischerweise auf eine mehr oder weniger große Menge an Kategorien festgelegt ist, wesentlich erweitert und es können semantische Bezüge zwischen visuellen Objekten und einer großen Menge an Wörtern oder Phrasen erfasst werden. Beispielsweise gibt es aktuelle Untersuchungen zur automatischen Erkennung eines inhaltlichen Bezugs zwischen Bildern und Textpassagen, (Hessel et al., 2019). Diese Art von Modellierung stellt einen Schritt in Richtung der Herstellung eines gemeinsamen Kontexts von Bild und Text dar. Für die multimodale Extraktion von Informationen aus fachwissenschaftlichen Texten ist es jedoch notwendig, solche referentiellen Beziehungen zwischen Text- und Bild-Teilen auch feingliedriger zu erfassen (Utescher and Zarriß, 2021).

Segmentierung und Objekterkennung: Die Grundlagen für die Rekonstruktion aus historischen Fotografien bilden die analytischen Verfahren der Photogrammetrie, d.h. die Gewinnung zweidimensionaler Bildinformationen (Wiedemann et al., 2000). Photogrammetrische Verfahren liefern auch räumliche Relationen zwischen Fotografien und dreidimensionalen Objektgeometrien. Aus durch bildgebende Verfahren entstandenen Datensätzen lassen sich einfache Strukturen (Vosselman et al., 2004), aber auch komplexe Objekte wie Gebäude (Li et al., 2016, Agarwal et al., 2011) automatisch segmentieren sowie zuordnen (Martinovic et al., 2015, Hackel et al., 2016). Dabei können auch Rückschlüsse darauf gezogen werden, welche Bildteile welche Teile der 3D-Objektgeo-

metrien referenzieren (Xie et al., 2016, Vosselman et al., 2004). Maschinelles Lernen (ML) spielt eine zunehmend größere Rolle bei der Segmentierung von Bild und der Objekterkennung (Minaee et al., 2021, Jiao et al., 2019) sowie der Strukturerkennung in 3D-Daten (Guo et al., 2020). Insbesondere vor dem Hintergrund der Anwendung von ML Ansätzen auf historische Quellen bestehen spezifische Herausforderungen in zumeist sehr kleinen und qualitativ heterogenen Ausgangsdaten (Fiorucci et al., 2020).

Verortung und Exploration: Ein besonderes Verhältnis besteht zwischen Bildquellen und 3D-Modellen. Im BMBF-Projekt *HistStadt4D* wurden exemplarisch Methoden zur automatischen Verortung von Fotografien entwickelt (Maiwald and Maas, 2021) und diese Daten in einer virtuellen 4D-Forschungsumgebung zugänglich gemacht (Bruschke et al., 2018, Maiwald et al., 2019) sowie quantitative raumbezogene Analysemethoden evaluiert (Dewitz et al., 2019). Mit Hilfe von Verfahren des maschinellen Lernens sollen in *HistKI* zudem Objektquellen und Textquellen (z.B. Bildunterschriften) verknüpft werden, um in Zukunft eine detaillierte Kontextualisierung und Verortung der Fotografien und Texte zu erlauben und damit über bisherige Methoden des *distant viewing* (Arnold and Tilton, 2019) hinauszugehen.

Forschungsaufriß

Im Folgenden soll ein kurzer Überblick über erste Forschungsschritte im Projekt gegeben werden.

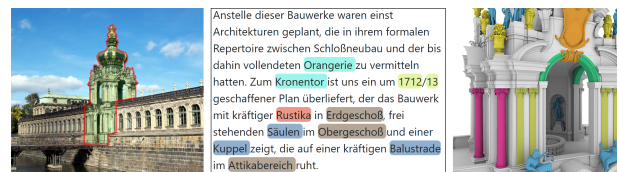


Abb. 1: Identifizierte Architekturelemente im Bild (links), im Text (mitte) und im 3D-Modell (rechts) am Beispiel des Kronentors des Dresdner Zwingers.

Identifikation von Forschungsszenarien

Der Einsatz von multimodalen KI-Techniken wird anhand von ausgewählten Szenarien untersucht, in denen Informationen aus Bildern, Texten und 3D-Modellen zur Beschreibung von Wissen über Architekturobjekte und städtebaulichen Ensembles für einen Analyseprozess miteinander verschrankt werden können. In Voruntersuchungen wurde dafür mit Hilfe qualitativer Expertenbefragungen und Workshops eine Reihe generischer Szenarien identifiziert (Kröber, 2021, Dewitz et al., 2019) und hinsichtlich einer Relevanz und Umsetzbarkeit priorisiert. Aus den ermittelten ca. 20 Szenarien wurden für eine erste Projektphase die medienübergreifende Identifikation von Objektbeschreibungen („Welche Bilder, Texte, 3D-Daten beschreiben dasselbe Objekt?“) sowie die Analyse von Beschreibungen (bspw.: „Wie kann die Datierung von historischen Bild- und Textdarstellungen von Objekten durch eine multimodale Validierung, bspw. anhand bereits datierter Medien, unterstützt werden?“) als Forschungsschwerpunkte ausgewählt.

Medienübergreifende Klassifikation

Die medienübergreifende Verarbeitung von 3D-Modellen, Bild- und Textquellen in einem möglichst universellen Format ist eine der zentralen Herausforderungen des Projekts. Eine diesbezüglich wichtige Voraussetzung ist es, medienübergreifend Elemente zu identifizieren und zu benennen. Die Entwicklung von domänenspezifischen Ontologien für architekturgeschichtliche Inhalte ist aktuell ein Schwerpunktthema und wird von einer Vielzahl von Initiativen vorangebracht – beispielhaft seien hier ICONCLASS¹, Getty Art & Architecture Thesaurus (AAT), als generische Ontologie Wikibase² sowie als übergreifende Referenzontologie CIDOC-CRM benannt. Als Grundgerüst für die strukturierende Beschreibung von Architekturelementen dient in unserem Projekt das AAT, welches trotz aktuell noch vorhandener Defizite hinsichtlich der deutschen Übersetzung durch eine umfassende und an einer kunsthistorischen Typisierung orientierte Struktur eine gute Ausgangsbasis für unser Projekt verspricht. Dabei wird von uns zunächst nur die gemessen am Gesamtvokabular kleine Untergruppe *architectural elements*³ verwendet. Die identifizierten Elemente im Text (einzelne Wörter oder Wortgruppen), Bild (polygonale Bildausschnitte) und 3D-Modell (einzelne Teilgruppenobjekte) werden den Konzepten des AAT zugeordnet (Abb. 1). Je nach Quellentyp sind dabei verschiedene Verfahren, wie z.B. semantic segmentation, named entity recognition (NER), und discourse parsing, notwendig, sowohl was die Identifizierung der Konzepte als auch die semantische Anreicherung betrifft. Als Nächstes müssen die identifizierten Konzepte zwischen den verschiedenen Quellen in einen Zusammenhang gebracht werden. Dazu werden die Identifier der AAT-Konzepte abgeglichen sowie meta-klassifiziert (z.B. Ionische Säule → Säule). Um mehrere Instanzen einer solchen Klasse zu unterscheiden (bspw. einzelne Säulen in einer Säulenreihe) werden die einzelnen Instanzen mit einer Identifikationsnummer versehen. Das AAT liefert ein detailliertes Vokabular zur Klassifizierung, reicht aber derzeit nicht aus, um alle Informationen, insbesondere Informationen aus Textquellen, vollständig abzubilden. Es müssen daher weitere Merkmale identifiziert werden, die das AAT nicht repräsentieren kann, wie z.B. Lagebeziehungen („Ostflügel“, „Westfassade“) oder Eigennamen.

Multimodale Datenanreicherung

In einem weiteren Schritt werden Ansätze zur multimodalen Datenanreicherung und -validierung entwickelt. So werden beispielsweise im 3D-Raum in Relation zum 3D-Modell verortete Bilder verwendet, um eine im 3D-Modell vorliegende Strukturierung auf Bilddokumente zu übertragen (Niebling et al., 2018). Durch den räumlichen Zusammenhang zwischen Bild und Modell können Annotationen von bereits semantisch angereicherten 3D-Modellen bzw. ihrer einzelnen Bauteile auf die entsprechenden Bildquellen projiziert, sowie auch von bereits annotierten Bildquellen auf 3D-Modelle zurückübertragen werden. Der 3D-Raum bietet zudem erweiterte Möglichkeiten, Lagebeziehungen zwischen (Teil-)Objekten herauszufinden. Die erkannten Zusammenhänge und semantischen Beziehungen werden in einer Ontologie gespeichert.

Ausblick

Quo vadis? Während im aktuellen Projektstadium mit manuell klassifizierten Quellen gearbeitet wird, ist ein nächster Schritt die Untersuchung von Ansätzen zur automatisierten Erkennung und Annotation von Objektbestandteilen. Hier werden schwerpunktmäßig KI-basierte Modelle verwendet, die auf die jeweiligen Modalitäten (3D-Modelle, Bilder, (multimodale) Texte) spezialisiert sind – beispielsweise die bereits benannten computerlinguistischen Verfahren zur Texterkennung sowie ein modulares Object Retrieval für die Erkennung von architektonischen Strukturen in Bildern (Münster et al., in print) und die in Schritt 3 ausgeführte Übertragung dieser Segmentierung auf 3D-Modelle.

Darauf aufbauend erfolgen in weiteren zukünftigen Schritten einerseits die Verbesserung und multimodale Validierung von Erkennungsqualitäten sowie die Entwicklung eines Demonstrators zur Nutzererprobung mit Historiker*innen.

Dies dient auch als Grundlage einer Bewertung von KI-Ansätzen für die historische Forschung. Hier gilt es beispielsweise, die Diskrepanz zwischen dem großen Datenbedarf der KI-Modelle und der Komplexität des geschichtswissenschaftlichen Expertenwissens zu untersuchen und damit zu bewerten, wie effektiv existierende KI-Modelle mit begrenzten Datenmengen für Teilaspekte der (architektur-)geschichtlichen Quellenkritik eingesetzt werden können.

Fußnoten

1. <http://www.iconclass.org/rkd/61F/> sowie <http://www.iconclass.org/rkd/47/>, 15.07.2021.
2. <https://www.wikimedia.de/projects/wikibase/>, 15.07.2021
3. <http://vocab.getty.edu/aat/300000885>, 15.07.2021.

Bibliographie

- Agarwal, S./ Furukawa, Y./ Snaveley, N./ Simon, I./ Curless, B./ Seitz, S. M. / Szeliski, R.** (2011). Building rome in a day. *Communications of the ACM*, 54, 105.
- Arnold, T. / Tilton, L.** (2019). Distant viewing: analyzing large visual corpora. *Digital Scholarship in the Humanities*.
- Bell, P. / Ommer, B.** (2019). Computer Vision und Kunstgeschichte – Dialog zweier Bildwissenschaften. In: KUROC-ZYNSKI, P., BELL, P. & DIECKMANN, L. (eds.) *Digital Art History*. Heidelberg.
- Brieber, D./ Nadal, M./ Leder, H. / Rosenberg, R.** (2014). Art in Time and Space: Context Modulates the Relation between Art Experience and Viewing Time. *PLoS ONE*, 9, e99019.
- Bruschke, J./ Maiwald, F./ Münster, S. / Niebling, F.** (2018). Browsing and Experiencing Repositories of Spatially Oriented Historic Photographic Images. *Studies in Digital Heritage*, 2, 138-149.
- Dewitz, L./ Kröber, C./ Messemer, H./ Maiwald, F./ Münster, S./ Bruschke, J. / Niebling, F.** (2019). HISTORICAL PHOTOS AND VISUALIZATIONS: POTENTIAL FOR RESEARCH. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W15, 405-412.
- Fiorucci, M./ Khoroshiltseva, M./ Pontil, M./ Travaglia, A./ Del Bue, A. / James, S.** (2020). Machine Learning for Cultural Heritage: A Survey. *Pattern Recognition Letters*, 133, 102-108.

- Fleck, L.** (1980). *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv*, Frankfurt a. M., Suhrkamp.
- Guo, Y./ Wang, H./ Hu, Q./ Liu, H./ Liu, L. / Bennamoun, M.** (2020). Deep Learning for 3D Point Clouds: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1.
- Hackel, T./ Wegner, J. D. / Schindler, K.** (2016). Fast semantic segmentation of 3D point clouds with strongly varying density. *ISPRS Annals*, 3, 177–184.
- Hessel, J./ Lee, L. / Mimno, D.** (2019). Unsupervised Discovery of Multimodal Links in Multi-image, Multi-sentence Documents. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.
- Holenstein, E.** (1988). *Linguistik, Semiotik, Hermeneutik*, Frankfurt am Main.
- Hoppe, S./ Breitling, S.** (2016). Virtual Palaces, Digital Images – an Introduction. In: S., H. & S., B. (eds.) *Virtual Palaces, Part II: Lost Palaces and their Afterlife. Virtual Reconstruction between Science and Media*. Heidelberg: arthistoricum.net.
- Jiao, L./ Zhang, F./ Liu, F./ Yang, S./ Li, L./ Feng, Z. / Qu, R.** (2019). A Survey of Deep Learning-Based Object Detection. *IEEE Access*, 7, 128837-128868.
- Kamposiori, C. / Benardou, A.** (2011). Collaboration in Art Historical Research: Looking at primitives. *Z. Kunstgeschichte* [Online]. Available at <http://www.kunstgeschichte-ejournal.net/157/> (accessed 17 June 2011).
- Klinke, H.** (2016). Big Image Data within the Big Picture of Art History. *Int. J. Digital Art History*, 2.
- Knorr-Cetina, K. / Reichmann, W.** (2015). Epistemic Cultures. *International Encyclopedia of the Social & Behavioral Sciences*. Amsterdam: Elsevier.
- Kohle, H.** (2018). *Initiative zur Einrichtung eines Schwerpunktprogramms. Das digitale Bild*.
- Köhler, T./ Günther, F./ Herbst, S./ Münster, S. / Fischer, H.** (2016). *Abschlussbericht im Projekt SUFES. Unterstützungsangebote und Strukturen im Themenfeld eScience*, Dresden.
- Kröber, C.** (2021). German Art History Students' use of Digital Repositories: an Insight *Papers Proceedings, Diversity, Divergence, Dialogue*. Cham: Springer LNCS.
- Kuroczynski, P./ Bell, P. / Dieckmann, L.** (eds.) 2019. *Digital Art History*, Heidelberg.
- Lee, E. / Münster, S.** (2018). Fishing for Knowledge in a Sea of Data (Session). *24th Annual Meeting of the European Association of Archaeologists*, 2018 Barcelona.
- Li, M./ Nan, L./ Smith, N. / Wonka, P.** (2016). Reconstructing building mass models from UAV images. *Computers & Graphics*, 54, 84-93.
- Maiwald, F./ Henze, F./ Bruschke, J. / Niebling, F.** (2019). Geo-Information Technologies for a Multimodal Access on Historical Photographs and Maps for Research and Communication in Urban History. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W11, 763-769.
- Maiwald, F. / Maas, H.-G.** (2021). An automatic workflow for orientation of historical images with large radiometric and geometric differences. *The Photogrammetric Record*, 36, 77-103.
- Martinovic, A./ Knopp, J./ Riemenschneider, H. / Van Gool, L.** (2015). 3d all the way: Semantic segmentation of urban scenes from start to end in 3d. *IEEE Computer Vision & Pattern Recognition*. 4456–4465.
- Minace, S./ Boykov, Y. Y./ Porikli, F./ Plaza, A. J./ Kehtarnavaz, N. / Terzopoulos, D.** (2021). Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1.
- Münster, S./ Apollonio, F./ Bell, P./ Kuroczynski, P./ Lenardo, I. D./ Rinaudo, F. / Tamborrino, R.** (2019). Digital Heritage meets Digital Humanities. *ISPRS Archives*, XLII-2/W15, 813–820.
- Münster, S./ Bruschke, J./ Maiwald, F. / Kleiner, C.** (in print). Software and content design of a browser-based mobile 4D VR application to explore historical city architecture. *Proceedings of the 3rd Workshop on Structuring and Understanding of Multimedia heritAge Contents*.
- Münster, S./ Kamposiori, C./ Friedrichs, K. / Kröber, C.** (2018). Image Libraries and their Scholarly Use in the Field of Art and Architectural History. *Int. J. Digital Libraries*, 19, 367–383.
- Münster, S. / Terras, M.** (2020). The visual side of digital humanities. A survey on topics, researchers and epistemic cultures in visual digital humanities. *Digital Scholarship in the Humanities*, 35, 366–389.
- Opgenoorth, E.** (1997). *Einführung in das Studium der neueren Geschichte*, Paderborn.
- Panofsky, E.** (1939). *Studies in Iconology. Humanistic Themes in the Art of the Renaissance*, Oxford, Oxford University Press.
- Polanyi, M.** (1966). *The tacit dimension*, Chicago, University of Chicago Press.
- Popper, K.** (1998). *Objektive Erkenntnis. Ein evolutionärer Entwurf*, Hamburg, Hoffmann und Campe.
- Precht, P.** (2002). *Edmund Husserl zur Einführung*, Hamburg.
- Reich, K.** (2006). *Konstruktivistische Ansätze in den Sozial- und Kulturwissenschaften. Konstruktivistische Didaktik: Lehr- und Studienbuch mit Methodenpool*. Beltz.
- Utescher, R. / Zarriß, S.** (2021). What Did This Castle Look like before? Exploring Referential Relations in Naturally Occurring Multimodal Texts. In *Proceedings of the Third Workshop on Beyond Vision and LAnguage: inTEgrating Real-world kNowledge (LANTERN)*.
- Vosselman, G./ Gorte, B. G./ Sithole, G. / Rabbani, T.** (2004). Recognising structure in laser scanner point clouds. *ISPRS Archives*, 46, 33-38.
- Wengenroth, U.** (1998). *Was ist Technikgeschichte?*, o. Ort.
- Wiedemann, A./ Hemmleb, M. / Albertz, J.** (2000). Reconstruction of historical buildings based on images from the Meydenbauer archives. *ISPRS Archives*, XXXIII, 887–893.
- Xie, J./ Kiefel, M./ Sun, M. T. / Geiger, A.** (2016). Semantic instance annotation of street scenes by 3d to 2d label transfer. *IEEE Computer Vision & Pattern Recognition*. 3688-3697.
- Ying, S./ Münster, S./ Köhler, T. / Sommer, C.** (unpublished). A look at the research on design ideas generation in industrial design: Review of literature from 2007 to 2016. *Int. J. of Design Creativity & Innovation*.
- Zöllner, H.-B.** (2005). *Hermeneutischer Zirkel und hermeneutische Differenz Perspektivität und Objektivität*.