

Der ferne Blick. Bildkorpora und Computer Vision in den Geistes- und Kulturwissenschaften - Stand - Visionen - Implikationen

Donig, Simon

simon.donig@uni-passau.de
Universität Passau, Deutschland

Handschuh, Siegfried

Siegfried.Handschuh@uni-passau.de
Universität Passau, Deutschland

Radisch, Erik

erik.radisch@uni-passau.de
Universität Passau, Deutschland

Rehbein, Malte

malte.rehbein@uni-passau.de
Universität Passau, Deutschland

Hastik, Canan

hastik@linglit.tu-darmstadt.de
Technische Universität Darmstadt, Deutschland

Kohle, Hubertus

hubertus.kohle@gmail.com
Ludwig-Maximilians-Universität München, Deutschland

Ommer, Björn

ommer@uni-heidelberg.de
Ruprecht-Karls-Universität Heidelberg, Deutschland

Ausgangslage

Getrieben von einer fortschreitenden Computerisierung ihrer wissenschaftlichen Methodik und einer wachsenden Verfügbarkeit von digitalen Bildwerken ist in den letzten Dekaden in zahlreichen Disziplinen der Geistes- und Kulturwissenschaften ein erneuertes Interesse am Bild als Forschungsgegenstand und Erkenntnisinstrument zu beobachten.

So begreift sich etwa die (digitale) Kunstgeschichte in der Erweiterung explizit als digitale Bildwissenschaft

(Kohle 2013), hat die Geschichtswissenschaft als “Visual History” das lange vernachlässigte Medium Bild als Quellengattung neben dem Text wiederentdeckt (Paul 2014), erfindet sich die Architekturgeschichte nicht zuletzt mit digitalen Instrumenten als Disziplin neu und haben sich junge Fächer wie die Material-Culture-Studies oder die Designgeschichte etabliert.

Alle diese Disziplinen begegnen sich in den Digital Humanities als einem geteilten Wissensraum, dessen Erkundung ganz wesentlich von den Möglichkeiten der Informatik mitbestimmt wird.

Die automatisierte Erschließung großer und größter Bildkorpora im Sinne eines effektiven Information Retrieval prägt dabei nach Jeffrey Schnapp die erste Welle der Digitalen Revolution (Schnapp 2011). Die zweite Welle - so wiederum Schnapp - müsse nun “qualitativ, interpretativ, erfahrungsbewusst, intuitiv erfassbar und schöpferisch” sein, um digitale Instrumente in den Dienst der Kernkompetenzen der Geistes- und Kulturwissenschaften stellen zu können: Achtsamkeit gegenüber Komplexität, Medienspezifika, historischem Kontext, analytischer Tiefe, Kritik und Interpretation.

Ziele des Panels

Mit dem hier vorgeschlagenen Panel verorten wir den Stand der digitalen Bildforschung in dieser idealtypischen Abfolge. Forschende aus unterschiedlichen Disziplinen beleuchten den State-of-the-art im Bereich automatisierter Verfahren der Computer Vision sowie gegenwärtige Wege und Visionen ihrer zukünftigen Anwendung in den Geistes- und Kulturwissenschaften. Wir erkunden Herausforderungen, die neue Werkzeuge wie die automatisierte Bildanalyse an Instrumentenkritik und Methodentransparenz stellen und tragen zur Konkretisierung von Best Practices in diesem Bereich bei.

Gegenstand

Verfahren, Praktiken und Projekte

Von Merkmalerkennung (Pattern Matching) bis hin zu Ansätzen aus dem Bereich des maschinellen Lernens wie etwa Deep Learning haben Verfahren der Computer Vision in den letzten Jahren das Potential entwickelt, sich als disruptive Technologie in den Geistes- und Kulturwissenschaften zu erweisen. Dieser Umstand verdankt sich unter anderem raschen Fortschritten bei der verfügbaren Rechenleistung, aber auch einem zunehmend vereinfachten Zugang zu geeigneter Software sowie einer wachsenden Verfügbarkeit digitaler Bilddaten.

Im Rahmen des Panels fragen wir zunächst nach den sich rasch entwickelnden Instrumenten und werfen einen Blick auf Bereiche und Projekte in denen diese zur Anwendung kommen. Wir wollen erkunden, welche

Möglichkeiten computerbasierte automatisierte Verfahren etwa der Objekt- und Merkmalerkennung für die Digitalen Geisteswissenschaften eröffnen (zu einer einführenden Diskussion transdisziplinärer Potentiale von Computer Vision und Kunstgeschichte vgl. (Bell & Ommer (2015) sowie dies. (2016)).

Zwischen datengetriebenen und interpretativ hermeneutischen Zugängen

Wir explorieren, ob und wie die anfangs postulierte Dichotomie von datengetriebenen und interpretativ-hermeneutischen Zugängen zur Generierung von Wissen (Drucker 2011) fortbesteht oder einem differenzierteren Modell Raum gegeben hat. Inwieweit lässt sich der am „Close Viewing“ geschulte Methodenapparat direkt auf Verfahren eines „Distant-Viewing“ (oder „Distant-Watching“ im Fall von Bewegtbildern) übertragen, oder inwieweit wird hier eine Wissenschaftstheorie des Digitalen noch zu entwickeln sein? Dies schließt die Frage einer adäquaten Instrumentenkritik ebenso ein, wie die nach Selektionsverfahren, Korpusbildung oder den Auswirkungen der Freigabepraxis von Bildmaterial durch Kulturinstitutionen (GLAM).

Die semantische Lücke, Annotation, Multimodalität

Die Analyse von „Big Image Data“ auch im Bereich der Forschung wirft die Frage auf, wie Forschungsansätze in den letzten Jahren versucht haben, die „semantische Lücke“ (Semantic Gap) im Umgang mit diesem Datenmaterial zu schließen.

Welche Rolle kommt etwa Normdaten (GND), Thesauri (ICONCLASS, AAT oder ULAN) oder Ontologien, die spezifisch für die Dokumentation kultureller Artefakte entwickelt wurden (CIDOC-CRM) bzw. Domäneontologien wie Neoclassica (Donig, Christoforaki, Handschuh 2016) zu?

Welche Rolle könnte alternativ die semantische Annotation (Oren, Möller, Scerri, Handschuh, & Sintek 2006) durch Zugänge etwa aus dem Bereich der Gamification und des Crowdsourcing wie in ARTIGO (Wieser, Bry, Bérard, Lagrange 2013) bieten?

Kann die Informatik hier mit neuen Verfahren zur Verknüpfung verschiedener digitaler Analyseansätze in multimodalen Artefakten – also etwa die Untersuchung der Verbindung von Bild und Text oder Bewegtbild und Ton – digitales Sehen ergänzen und verbessern? (Bruni, Tran, Baroni 2014), (Hiippala 2013).

Methodentransparenz & Best Practices

In der Tat wirft ein digitales Sehen, analog zu kritischen Debatten um „Close“ und „Distant Reading“ (Bonfiglioli

& Nanni 2015) grundlegende epistemologische Fragen nach den Gemeinsamkeiten oder Unterschieden zwischen computergestützter Bildanalyse und in spezifischen Praktiken des Sehens geschulten Menschen auf. Erweitert etwa der ferne Blick nur die Geschwindigkeit und Menge dessen, was wir wahrnehmen können, oder verändert er unsere Wahrnehmung selbst?

Wir erkunden wie digitale Instrumente beschaffen sein müssen, um den Erfordernissen der Geistes- und Kulturwissenschaften nach Methodentransparenz zu genügen. (Für eine grundsätzliche Diskussion der Konsequenzen des Designs von Klassifizierungs- und Rankingmechanismen sowie verschiedene Formen von Opacity siehe Burell 2016). Wie kann etwa der Black Box Charakter von Neuronalen Netzen sinnvoll akkommodiert werden? Nicht zuletzt stellt sich die Frage nach dem Zusammenspiel von Instrument und Forschungspraxis. Reproduziert etwa ein (in sich vollständig transparenter) Klassifikator, der aber auf der Basis eines bestimmten Korpus trainiert worden ist, nicht letztlich spezifische kulturelle Konzepte und konfirmiert damit einen bestehenden Kanon?

Wir wollen uns diesen Fragen nicht verschließen, aber sie gleichermaßen rekontextualisieren wie über ihre forschungspraktischen Implikationen nachdenken. Wir werden dabei gerade mit jenen, die an der Weiterentwicklung dieser Instrumente arbeiten, diskutieren, wie Technologien gebaut und eingesetzt werden können, um ein emanzipatorisches und exploratives Potential zu entfalten.

In diesem Sinn wollen wir abschließend Ideen für Best Practices im Sinne eines methodisch angeleiteten Umgangs mit großen Bildkorpora zusammentragen, indem wir etwa nach der Rolle und Modellierbarkeit von Kontextualisierung oder der Rolle von Ontologien und Normdaten dabei fragen.

Impulsvorträge

Siegfried Handschuh, Universität Passau: Bild, multimodaler Verbund und Automatisierung

Bildanalyse stellt eine besondere Herausforderung an die Fähigkeit von computerbasierten Verfahren Wissen zu generieren. Dass Bilder vollkommen ohne einen Bezug zu anderen Modi – hier besonders Text – vorkommen ist eher eine Ausnahme als die Regel. Der multimodale Verbund von Bild und Text kann so durch Verfahren aus dem Bereich des maschinellen Lernens ausgewertet werden, um Wissen zu erzeugen, das einen breiteren Kontext berücksichtigt. Das Impulsstatement thematisiert diesen Zusammenhang insbesondere eine Methodik für die Repräsentation, Annotation und (teil)automatisierte Entdeckung multimodaler Wissensbestände in großen digitalen Materialkorpora und fragt wie Techniken der Computer Vision gewinnbringend mit Verfahren aus

dem Bereich der Verarbeitung Natürlicher Sprache zusammengebracht werden können.

Canan Hastik, Technische Universität Darmstadt:
Wie viel Semantik ist nötig und wie viel Automatisierung ist möglich?

Computerbasierte Verfahren sind ein wichtiges Instrument, um die digitale Bildforschung effizienter zu gestalten und zu verstetigen. Es wäre jedoch ein großer Fehler, Verfahren des maschinellen Lernens lediglich dafür einzusetzen, bestehende theoretische und historische Kategorien zu definieren und die Automatisierung für Digital Humanities Experten voranzutreiben. Die Herausforderung besteht darin, den DH Forscher bei der Kuration und Entwicklung von aussagekräftigen, repräsentativen und authentischen Forschungskorpora zu unterstützen. Ontologien sind dabei elementar für die Semantifizierung von großen bildbasierten Korpora und unterstützen zudem die kollaborative Entwicklung und die Verknüpfung von Wissensbeständen. Automatisierte Verfahren gilt es insbesondere dafür einzusetzen, zeitgenössische Kultur neu zu sehen und zu interpretieren.

Hubertus Kohle, Ludwig-Maximilians-Universität München: Ähnlichkeitsbestimmung in der digitalen Kunstgeschichte

Das Statement widmet sich aus einer eher geisteswissenschaftlichen Perspektive heraus der digital gestützten Ähnlichkeitsbestimmung. Ähnlichkeit wird hier verstanden als Erkenntnismotor, der insbesondere in der Kunstgeschichte seinen Ort hat, die schon immer mit dem vergleichenden Sehen eine Methode des Erkenntnisgewinns besaß. Abzuwägen wird sein, ob hierfür eher eine Metadatenanalyse oder die direkte Bildadressierung oder eine Mischung aus beidem zielführend ist.

Björn Ommer, HCI/IWR Ruprecht-Karls-Universität Heidelberg:

Aktuelle Entwicklungen im Bereich des Maschinellen Lernens und der Computer Vision haben Algorithmen hervorgebracht, die Visual Retrieval mit einer zuvor ungeahnten Performanz ermöglichen. Dadurch ergeben sich für die digitalen Bildwissenschaften ganz neue Möglichkeiten große Korpora zu erschließen und zu analysieren. Allerdings bringen diese neuen informatischen Verfahren auch ihre ganz eigenen Beschränkungen mit sich, die insbesondere bei ihrer Rekontextualisierung in den Geisteswissenschaften zutage treten. Dieser Vortrag wird das Potential einer interdisziplinären Kooperation von Informationswissenschaften und den Geisteswissenschaften diskutieren, grundlegende

Beschränkungen beleuchten und mögliche Auswege präsentieren.

Malte Rehbein, Universität Passau: Moderation

Bibliographie

Bell, Peter / Ommer, Björn: Training Argus, Ansätze zum automatischen Sehen in der Kunstgeschichte, in: *Kunstchronik* 68, Nr. 8 (2015): 414–20.

Bell, Peter / Ommer, Björn: Digital Connoisseur? How Computer Vision Supports Art History, in: Aggularo, A. & Albl, S. (eds.): *Il metodo del conoscitore approcci, limiti, prospettive - Connoisseurship nel XXI secolo*. Artemide, Roma 2016: 187-200.

Burrell, Jenna: How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms, in: *Big Data & Society* 3, Nr. 1 (2016): 1-12. doi:10.1177/2053951715622512.

Bonfiglioli, Rudi / Nanni, Federico: From Close to Distant and Back: How to Read with the Help of Machines, in: Gadducci, Fabio / Tamosanis, Mirko (eds.): *History and Philosophy of Computing*. IFIP Advances in Information and Communication Technology. Springer, Cham, 2015: 87–100. doi:10.1007/978-3-319-47286-7_6.

Bruni, Elia / Tran, Nam Khanh / Baroni, Marco: Multimodal Distributional Semantics, in: *Journal of Artificial Intelligence Research* 49, Nr. 1 (2014): 1–47.

Donig, Simon / Christoforaki, Maria / Handschuh, Siegfried: Neoclassica - A Multilingual Domain Ontology. Representing Material Culture from the Era of Classicism in the Semantic Web, in: Bozic, B. / Mendel-Gleason, G. / Debruyne, C. / O'Sullivan, D. (eds.): *Computational History and Data-Driven Humanities. CHDDH 2016*. Cham: Springer, 2016: 41–53. https://link.springer.com/chapter/10.1007/978-3-319-46224-0_5.

Drucker, Johanna: Humanities approaches to graphical display, in: *Digital Humanities Quarterly* 5, Nr. 1 (2011): 1–21.

Hastik, Canan / Steinmetz, Arnd / Thull, Bernhard: Using CIDOC CRM for Audiovisual Art, in: Smite, R. / Manovich, L. / Smits, R. (eds.): *Data Drift. Archiving Media and Data Art in the 21st Century*. Riga: RIXC 2015: 59-71.

Hiippala, Tuomo: The Interface between Rhetoric and Layout in Multimodal Artefacts, in: *Literary and Linguistic Computing* 28, Nr. 3 (2013): 461–71. doi:10.1093/llc/fqs064.

Kohle, Hubertus: *Digitale Bildwissenschaft*. Boizenberg: Hülsbusch, Werner, 2013.

Manovich, Lev: Data Science and Digital Art History. In: *International Journal for Digital Art History*, n. 1, 2015: 12-35. doi:hp://dx.doi.org/10.11588/dah.2015.1.21631., 30.

Oren, Eyal / Möller, Knud / Scerri, Simon / Handschuh, Siegfried / Sintek, Michael: What are Semantic Annotations. Relatório Técnico. DERI Galway 9, 62 (2006).

Paul, Gerhard: „Visual History“, in: *Docupedia-Zeitgeschichte* (13. März 2014). doi:<http://dx.doi.org/10.14765/zzf.dok.2.558.v3>.

Wieser, Christoph / Bry, François / Bérard, Alexandre / Lagrange Richard: ARTigo: building an artwork search engine with games and higher-order latent semantic analysis, in: *First AAAI Conference on Human Computation and Crowdsourcing*, 2013. <http://www.aaai.org/ocs/index.php/HCOMP/HCOMP13/paper/view/7634>.