

Den Menschen als Zeichen lesen. Quantitative Lesarten körperlicher Zeichenhaftigkeit in visuellen Medien

Howanitz, Gernot

gernot.howanitz@uni-passau.de
Universität Passau, Deutschland

Radisch, Erik

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

Decker, Jan-Oliver

Jan-Oliver.Decker@Uni-Passau.De
Universität Passau, Deutschland

Rehbein, Malte

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

Trotz der Textlastigkeit der Digital Humanities ist in der letzten Zeit ein gewisser beginnender „visualistic turn“ in der Disziplin zu beobachten, der in den Kulturwissenschaften bereits seit einiger Zeit konstatiert wird. (Sachs-Hombach 1993) Auch der Schwerpunkt dieser Konferenz weist in diese Richtung. Visuelle Medien bieten zweifelsohne eine große Chance für eine Weiterentwicklung der Digital Humanities, trotzdem ist festzustellen, dass sich dieser Turn bisher noch größtenteils auf die Kunstgeschichte konzentriert.¹ Das liegt aufgrund des visuellen Schwerpunktes dieser Disziplin zwar nahe, jedoch widmen sich auch andere Disziplinen, in unserem Falle die Kulturwissenschaften, aus der Perspektive ihrer spezifischen Fragestellungen den visuellen Medien.

Das Anliegen dieses Vortrages ist es, einen Zugang zu präsentieren, der es Kulturwissenschaftler/innen ermöglicht, große Bild- und Videokorpora in einem Methodenmix sowohl qualitativer als auch quantitativer Verfahren zu erfassen. Auf der letzten DHd-Konferenz haben wir erste Vorarbeiten in Form eines Ansatzes präsentiert, der aus visuellen Medien den relevanten symbolischen Kontext identifiziert und damit eine automatisierte Informationsextraktion und -analyse ermöglicht (Bermeiteinger/Howanitz/Radisch 2018). Ein großes Videokorpus konnte auf diese Weise bearbeitet werden, ohne alle Videos im einzelnen ansehen zu müssen. Das Auftreten von Menschen in diesen Videos oder Bildern wurde dabei lediglich am Rande geschnitten, indem

die Möglichkeit in Erwägung gezogen wurde, bestimmte Personen auf Einzelframes zu identifizieren.

Hier möchte unser diesjähriger Vortrag ansetzen und Jurij Lotmans (1984) semiotischen Ansätzen folgend menschliche (Ab-)Bilder zeichenhaft in visuellen Korpora lesen. Als Beispielkorpus dient uns eine Sammlung von Starpostkarten von Marlene Dietrich aus den 1930er und 1940er Jahren. Ziel ist es, ähnlich wie bei Distant Reading-Ansätzen zu versuchen, versteckte Muster, Gemeinsamkeiten und Abweichungen ohne Betrachtung der Einzelbilder identifizieren zu können. Dabei erweist sich das Erscheinen von Menschen in diesen Bildern als zentral und muss besonders berücksichtigt werden. Um menschliche Abbildungen in visuellen Korpora möglichst umfassend analysieren zu können, schlagen wir eine dreistufige Herangehensweise vor, nämlich die Identifikation einzelner Personen, die Analyse von Körperhaltungen, sowie die Mimikereffassung. Im Folgenden werden alle drei Teilbereiche vorgestellt.

Identifikation

In vielen Fällen der Analyse menschlicher Zeichenhaftigkeit ist es von großem Vorteil, bestimmte Personen identifizieren zu können. So ist es zum Beispiel interessant, die Auswahl an Personen in einem visuellen Korpus durch Kriterien der Relevanz in Bezug auf eine konkrete Forschungsfrage einzuschränken. Auch können auf diese Weise verschiedene Protagonisten und in weiterer Folge deren Kookkurrenzen innerhalb des Korpus identifiziert werden, was Rückschlüsse auf Personenkonstellationen zulässt. Handelt es sich bei den analysierten Bildern um Einzelframes aus Videos, erlaubt es dieser Ansatz zudem, resultierende Muster der Personenkonstellationen als Netzwerke zu visualisieren – analog zu der von Franco Moretti etablierten Netzwerkanalyse von Dramen, die in den letzten Jahren zahlreiche Anwendung gefunden hat.

Zum Erkennen von Gesichtern verwenden wir David Sandbergs *Facenet*, das auf OpenFace (Amos/Ludwiczuk/Satyanarayanan 2016) beruht und folgen dabei den Empfehlungen unseres Kollegen Sebastian Gassner (Gassner 2018). Hinzuweisen ist aber hier ebenfalls auf das Distant-Viewing Toolkit von Lauren Tilton und Taylor Arnold, das ebenfalls ermöglicht, bestimmte Akteure in Filmen identifizieren zu können (Tilton/Arnold 2018).

Haltung

Mit der bloßen Anwesenheit bestimmter Personen in visuellen Medien ist aber noch kaum Aussagen über deren Zeichenhaftigkeit möglich. Vielmehr gilt es auch die Körperhaltung der Personen einer eingehenden Analyse zu unterziehen. Gerade hier manifestiert sich die Zeichenhaftigkeit eines Körpers besonders. Der gesamte Körper fungiert als ein Zeichen. Je nach

Haltung können völlig unterschiedliche Aussagen getätigt werden, die nur zum Teil einerseits bewusst intendiert sind, andererseits bewusst wahrgenommen werden. Hier erlauben es algorithmische Verfahren, einen Blick auf das Phänomen zeichenhafter Körperlichkeit zu ermöglichen, der etwas weniger durch menschliche Erfahrung beeinflusst ist als dies bei qualitativen Ansätzen der Fall ist.

Körperhaltungen lassen sich auf Grundlage einiger wichtiger Keypoints (Bourdev/Malik 2009) analysieren, die die Position der Hände, Ellenbogen, Schultern, der Hüfte, der Knie und der Füße sowie des Kopfes angeben. Je nachdem, wie diese Positionen in Relation zueinander stehen, ergeben sich verschiedene Haltungen die Rückschlüsse auf symbolische, über den Körper kommunizierte Bedeutungen erlauben. Solche Ansätze wurden erst kürzlich von Leonardo Impetti und Franco Moretti zur Klassifizierung von Aby Warburgs Bilderatlas verfolgt (Impetti/Moretti 2017). Ähnliche Ansätze werden auch in unserem Projekt herangezogen. So erlaubt es beispielsweise die Clusterung nach den Keypoints in unserem Korpus bestimmte Körperhaltungen zusammenzufassen und deren Entwicklung diachron über die Zeitspanne des Korpus zu untersuchen. Der Zusammenhang zwischen Körperhaltungen und Genderstereotypen ist vielschichtig (für einen körpersemiotischen Zugang vgl. Mühlens-Achs 1998); gerade im Zusammenhang mit Marlene Dietrichs androgyner Selbstinszenierung, die männlich konnotierte Elemente aufgreift, ermöglicht eine quantitative Analyse dieser Körperhaltungen diesbezüglich neue Einblicke. Daneben interessieren uns auch Kopfhaltung und insbesondere die Blickrichtung, die ihrerseits die Grundparameter der zugrunde liegenden kommunikativen Situation offenbart.

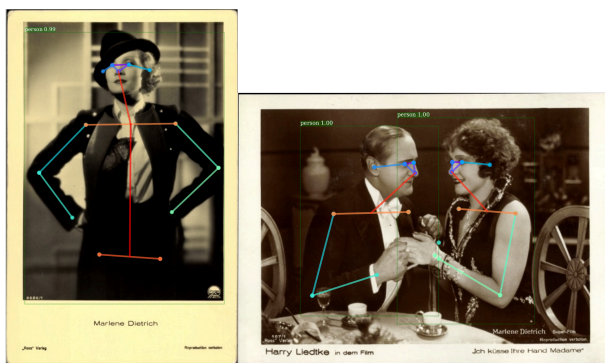


Abb. 1: Detectron identifiziert Marlene Dietrichs typische Doppelamphorenhaltung (links), rechts eine Figurenkonstellation.

Der Ansatz der Posen-Erkennung hat noch ein wesentlich höheres Potential für zukünftige Forschung. Hierüber könnten sich auch Bewegungsabläufe in Videos analysieren lassen, da sich bestimmte Bedeutungsebenen erst im Kontext des Zusammenspiels ergeben.

Technisch greifen wir einerseits auf die Keypoint-Ebene von *Detectron* (Girshik et al. 2018) zurück, eines Frameworks für Deep Learning und Objekterkennung, zur Verfügung gestellt von *Facebook Artificial Intelligence Research*, andererseits auf das von Cao et al. (2017) veröffentlichte System zur Posen-Erkennung, das im Vergleich zu *Detectron* weitere Keypoints identifiziert und darüber hinaus zwischen den beiden Körperhälften unterscheiden kann.

Ein Nachteil der hier verwendeten Algorithmen ist, dass sie die Keypoints nur zweidimensional erkennen, also die Position auf dem Bild erfassen, nicht aber deren Lage im Raum. Für eine dreidimensionale Erfassung zumindest des Gesichts wird mit *OpenFace* (Baltrušaitis et al. 2018) deshalb noch ein dritter Algorithmus verwendet, der eine Abschätzung der dreidimensionalen Lage – insbesondere auch der Kameraposition in Relation zum Gesicht – und der Blickrichtung erlaubt. Mithilfe dieser Informationen können grobe Rückschlüsse auf die Gesamtkomposition einer Szene gemacht werden.

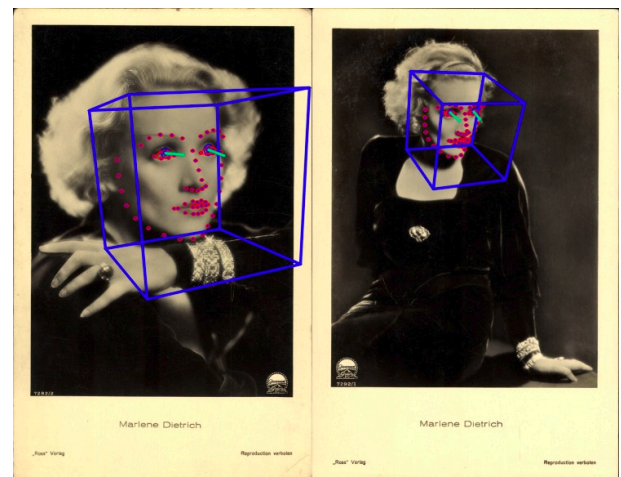


Abb. 2: OpenFace blickt Marlene Dietrich ins Gesicht: Lage im Raum (blau), Keypoints (rot), berechnete Blickrichtung (grün)

Mimik

OpenFace erlaubt es des Weiteren eine weitere Ebene menschlicher Zeichenhaftigkeit in visuellen Medien zu erschließen – das Gesicht. Dabei wird neben der dreidimensionalen Position des Kopfes ebenfalls alle wichtigen Punkte innerhalb des Gesichtes, darunter die Umrisse von Augen, Nase, Lippen, Augenbrauen und Kinn erkannt (Abb. 2). Der Algorithmus ermittelt die Abweichung dieser Gesichtspunkte von einem Standardmodell, welche wiederum mithilfe des Facial Action Coding System (FACS), ein System der Taxonomie menschlicher Gesichtsausdrücke, beschrieben werden können (Ekman/Friesen 1978). FACS besteht aus einer Reihe von Action Units, die basale Gesichtsbewegungen

klassifizieren, wie beispielsweise die Hebung der inneren Augenbrauen (AU1), die Senkung der Augenbrauen (AU4), das Zusammenkneifen der Lider (AU7) und das Anziehen der Mundwinkel (AU12).

Diese AUs stellen einen Standard zur Beschreibung von Gesichtsausdrücken dar.

Damit stellt OpenFace ein ideales Werkzeug dar, um Mimiken in visuellen Medien auslesen zu können, mit einer großen potenziellen Breite an Einsatzmöglichkeiten. So lassen sich beispielsweise damit Bilder mit Personen nach Gesichtsausdrücken clustern oder die Häufigkeit bestimmter Action Units im Zeitverlauf zu analysieren, um so Rückschlüsse über veränderte kulturelle Praxen machen zu können. Im Falle von Marlene Dietrich ist beispielsweise beobachtbar, dass AU2 (Hebung der äußeren Augenbrauen) eine Zeit lang häufig auftritt, was auf die Schminkgewohnheiten der damaligen Zeit zurückzuführen ist. Auf diese Weise können auch Modetrends und Stilfragen bis zu einem gewissen Grad quantitativ erfasst werden.

Solche Informationen können aber auch in anderen Kontexten von großem analytischen Wert sein. In einem Korpus von 290 Zigarettenbildern zeigen sich etwa klare Unterschiede zwischen den Geschlechtern. AU6 (Wangen heben) kombiniert mit AU12 (Mundwinkel heben) entspricht etwa der Emotion „Freude“, die bei den Frauen häufiger anzutreffen ist. AU15 (Mundwinkel senken) ist hingegen bei den Männern weitaus verbreiteter (siehe Abb. 3).

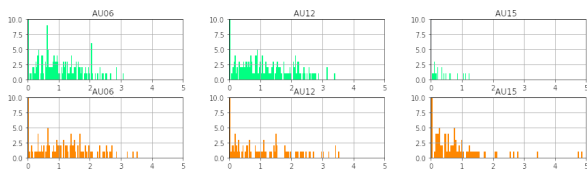


Abb. 3: Histogramme von AU6, AU12 und AU15 bei Frauen (grün) und Männern (orange).

Das Puzzle setzt sich zusammen: Quantitative Lesarten körperlicher Zeichenhaftigkeit

Die oben beschriebene Methodenkombination ist einerseits nötig, um körperliche Zeichenhaftigkeit in visuellen Medien auf den verschiedenen skizzierten Bedeutungsebenen erfassen zu können, erschwert aber andererseits das Arbeiten mit dem Korpus und verlangt deshalb nach einer entsprechenden methodischen Aufbereitung. Das erste Ziel unseres Beitrages ist deshalb, die Methoden durch ein Testkorpus von rund 200 Bildern systematisch zu evaluieren. Anschließend wenden wir die Methoden auf ein Korpus von mehreren tausend deutschen Starpostkarten vorwiegend aus den 1930ern

und 1940ern an, die wir in Zusammenarbeit mit dem Filmmuseum Potsdam digitalisiert haben. Als Basis unserer Arbeiten dient ein eigenständiges Framework, das wir für mehrere Forschungsprojekte innerhalb des vom BMBF geförderten Passau Centre for eHumanities (PACE, Fördernummer: 01UG1602) entwickelt haben und die es ermöglichen verschiedene quantitative Aspekte mit qualitativen Fragestellungen kurzzuschließen. Wie bereits erwähnt, ist es ein *desideratum*, das System auch für Bewegtbilder zu öffnen und weiterzuentwickeln, auch wenn der damit verbundene Aufwand nicht trivial ist.

Fußnoten

1. Der Schwerpunkt auf Kunstgeschichte spiegelt sich beispielsweise im DFG-Schwerpunktprogramm „Das digitale Bild“ wieder, dass zwar prinzipiell auch für die Kulturwissenschaften geöffnet war, aber in der Ausschreibung einen deutlichen Fokus auf die Kunstgeschichte nahm.

Bibliographie

Amos, Brandon / Ludwiczuk, Bartosz / Satyanarayanan, Mahadev (2016): *"Openface: A general-purpose face recognition library with mobile applications,"* CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016. <https://www.cs.cmu.edu/~satya/docdir/CMU-CS-16-118.pdf> [letzter Zugriff 29. 9. 2018].

Arnold, Taylor / Tilton, Lauren (2018): *„Distant Viewing Toolkit (DVT) for the Cultural Analysis of Moving Images“*, <https://github.com/distant-viewing/dvt> [letzter Zugriff 29. 9. 2018].

Baltrušaitis, Tadas / Zadeh, Amir / Chong Lim, Yao / Morency, Louis-Philippe (2018): *“OpenFace 2.0: Facial Behavior Analysis Toolkit“*, IEEE International Conference on Automatic Face and Gesture Recognition 2018., <https://ieeexplore.ieee.org/document/8373812> [letzter Zugriff 29. 9. 2018].

Bermeiter, Bernhard / Howanitz, Gernot / Radisch, Erik (2018): *“Contextualising Bandera. Eine Distant Watching-Methode“*, DHd 2018.

Bourdev, Lubomir / Malik, Jitendra (2009): *“Poselets: Body Part Detectors Trained Using 3D Human Pose Annotations“*, ICCV 2009. https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/human/poselets_iccv09.pdf [letzter Zugriff 29. 9. 2018].

Ekman, Paul / Friesen, Wallace V. (1978): *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting.

Gassner, Sebastian (2018): *„How to Use David Sandberg's Facenet Implementation“*, <https://github.com/sebastian/facenet/blob/master/HOWTO.md> [letzter Zugriff 29. 9. 2018].

Girshick, Ross / Radosavovic, Ilija / Gkioxari, Georgia / Dollár, Piotr / He, Kaiming (2018): *Detectron*. <https://github.com/facebookresearch/detectron> [letzter Zugriff 29. 9. 2018].

Cao, Zhe / Simon, Tomas / Wei, Shih-En / Sheikh, Yaser (2017): „*Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*“. CVPR 2017. <https://arxiv.org/pdf/1611.08050.pdf> [letzter Zugriff 29. 9. 2018].

I mpett, Leonardo / Moretti, Franco (2017): „*Totentanz. Operationalizing Aby Warburg's Pathosformeln*.“ In: Literary Lab Pamphlet 16, November 2017, <https://litlab.stanford.edu/LiteraryLabPamphlet16.pdf> [letzter Zugriff 29. 9. 2018].

Lotman, Jurij (1984): „*O semiosfere*“. U#en. zap. Tart. gos. un-ta 641 (1984), 5-23. (=Trudy po znakovym sistemam 17).

Mühlen-Achs, Gitta (1998): *Geschlecht bewusst gemacht. Körpersprachliche Inszenierungen*. München: Frauenoffensive.

Sachs-Hombach, Klaus (1993): *Das Bild als kommunikatives Medium. Elemente einer allgemeinen Bildwissenschaft*. Köln: Halem.