

Digitale Empirie?

Computergestützte Filmanalyse im Spannungsfeld
von Datenmodellen und Gestalttheorie

Jan-Hendrik Bakels / Matthias Grotkopp / Thomas Scherer /
Jasper Stratil

Digitale, das heißt software-gestützte und algorithmische Methoden der Filmanalyse haben in den letzten 15 Jahren Möglichkeiten eröffnet, audiovisuelle Bilder empirisch zu «vermessen». Diese methodologischen Innovationen werfen jedoch zugleich die Frage nach ihrem explikativen Potenzial – und dessen Grenzen – auf. Diese Frage erscheint umso gewichtiger, will man die Daten, welche derartige Tools produzieren, in Relation zu jener Dimension audiovisueller Bilder setzen, welche ihrer kulturellen Rezeption und Zirkulation zentral ist: dem Erleben der Zuschauenden. Setzt dies doch einen theoretischen Ansatz voraus, der dem Verhältnis von Mustern «messbarer» Basisdaten (z. B. timecode-basierte Angaben zu Schnitten, Kamerabewegungen, Musik oder Dialog-Passagen) zum Wahrnehmen, Fühlen und Verstehen der Zuschauenden konzeptionell Bedeutung beimisst. Christian Metz hat lange vor der Entwicklung digitaler Methoden der Filmanalyse die grundlegende Problematik eines (analytischen) Unterfangens, welches das Erleben und Verstehen filmischer Bilder adressiert, in einem der berühmtesten Bonmots der Filmtheorie auf den Punkt gebracht: «Gerade weil der Film leicht zu verstehen ist, ist er so schwer zu erklären» (1973, 101).

In der Geschichte der Filmtheorie als Geschichte der Erklärungsversuche dieses Verhältnisses ist die Relation von Film und Zuschauenden weniger als «kleinster gemeinsamer Nenner» zu fassen, sondern erweist sich vielmehr als eine dynamische Konstellation, über deren jeweilige Inter-

pretation sich die Theorieschulen voneinander abzugrenzen suchen. Gibt es auch eine Vielzahl unterschiedlicher filmtheoretischer Perspektiven, so kann doch die heuristische Annahme einer Spannweite zwischen Ansätzen, welche von einem Primat der nachvollziehenden Verstehensprozesse ausgehen, und jenen, welche die in der Bewegungsdimension des audiovisuellen Bildes fundierte, dynamische affektive Erfahrung zum Ausgangspunkt nehmen, eine erste Orientierung liefern.¹ Mag es auch verlockend sein, die Geschichte der Filmtheorie auf Grundlage dieser beiden vagen Theoriestränge in ein Narrativ von Opposition oder dialektischem Wechsel zu fassen, möchte der vorliegende Artikel vielmehr ihre Berührungspunkte zum Ausgangspunkt nehmen, um so Potenziale und Grenzen digital gestützter, formalästhetisch perspektivierter Analysemethoden innerhalb der Filmwissenschaft aufzuzeigen. Doch wo liegen diese Berührungspunkte? Und: Lässt sich ein solcher Berührungspunkt auf der Ebene systematisch gewonnener filmanalytischer Daten ausmachen?

Die erste zentrale These dieses Artikels gründet in der Annahme, *dass sich die audiovisuelle Komposition, d. h. die Art und Weise, wie Bilder und Töne in audiovisuellen Sequenzen konkret in der Zeit arrangiert sind, als eine formalästhetische Matrix sowohl von Prozessen des Verstehens als auch des Fühlens der Filmzuschauenden fassen lässt*. Bereits in den vielzähligen Schriften Sergej Eisensteins zur Theorie und Praxis der Montage wird diese mal als eine Anordnung von Zeichen und Symbolen gefasst (Eisenstein 1988), an anderer Stelle wiederum als strukturelle Grundlage kinästhetischer Phänomene (Eisenstein 2006a). An ein solches Verständnis der Montage lassen sich sowohl Theorien des Filmverstehens als auch Theorien der somatischen Erfahrung anschließen. Erstere konzeptualisieren die kompositorischen Muster als Grundlage bedeutungsgenerierender Informationsverarbeitung (vgl. Bordwell 1989), letztere rekonstruieren in den kompositorischen Mustern die Strukturierung eines verkörperten

1 Auch wenn eine entsprechende Verortung einzelner Ansätzen deren Argumentation im Detail natürlich nicht gerecht werden kann, so lassen sich doch exemplarisch für die Theorien des russischen Formalismus (vgl. Beilenhoff 2005), die semiotischen Ansätze des französischen Strukturalismus (Metz 1973) und den kognitivistisch begründeten Neo-Formalismus (vgl. Kirsten 2016) Fragen des Filmverstehens als eine gemeinsame Perspektive ausmachen. Der Zusammenhang zwischen Bewegung und Emotion bzw. Affekt hingegen steht im Zentrum etwa der Filmpsychologie Münsterbergs (1996), der späteren Schriften Sergej Eisensteins (2006a), der Filmphilosophie Gilles Deleuzes (1997) oder der neo-phänomenologischen Filmtheorie (Sobchack 1992) sowie im Ansatz der Cinematic Metaphor (Müller/Kappelhoff 2018).

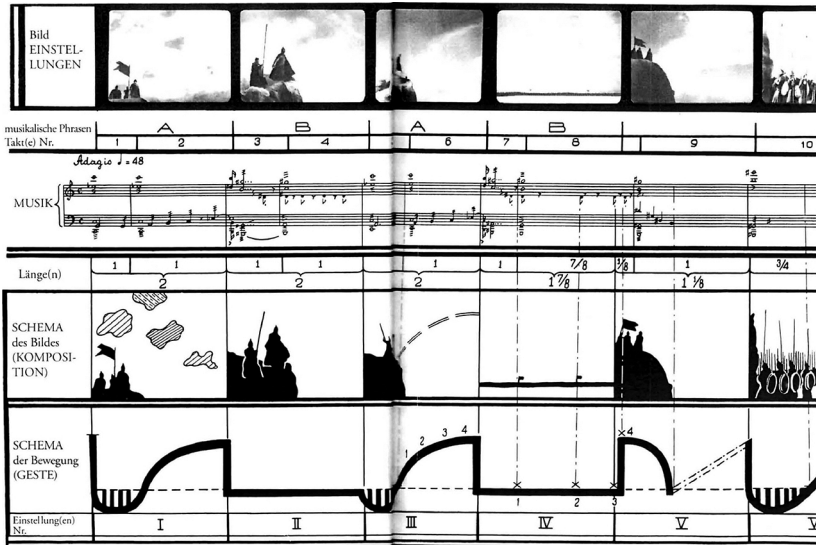
Wahrnehmungsprozesses (vgl. Sobchack 1992). Unsere zweite zentrale These trägt die erste ins Feld filmanalytischer Methodologie: *Lassen sich die Muster audiovisueller Komposition empirisch fassen und evident machen, so sollten sich darüber Einsichten in eben jene Prozesse des Filmverstehens und der affektiven Erfahrung filmischer Bilder ergeben.*

Doch um derartige empirische Studien filmischer Poetik und Ästhetik umzusetzen, bedarf es zunächst zweier wesentlicher Grundlagen: Zum einen müssen sinnfällige theoretische Prämissen in Anschlag gebracht werden, um die bloße Struktur solcher Muster auf das Erfahren, Fühlen und Verstehen filmischer Bilder beziehen zu können. Zum anderen bedarf es einer gleichermaßen komplexen wie systematischen analytischen Herangehensweise, um den Anspruch empirischer Genauigkeit sowie die Vergleichbarkeit derartiger Studien zu gewährleisten. Mit Letzterem verbindet sich schließlich zugleich die Notwendigkeit von geeigneten Werkzeugen, derartig aufwändige Detailstudien überhaupt in größerem Stil betreiben zu können – hier kommen die Möglichkeiten digitaler Filmanalyse ins Spiel.

Vom kompositorischen Muster zur Erfahrung der Zuschauenden

Im Folgenden möchten wir anhand der Ausdrucksbewegungstheorie (vgl. Kappelhoff/Bakels 2011) veranschaulichen, wie digital rekonstruierte kompositorische Muster audiovisueller Sequenzen im Hinblick auf Dynamiken der Zuschauendenerfahrung qualifizierbar sind.² Bereits zentrale Arbeiten der frühen Filmtheorie (Münsterberg 1996; Balázs 2001) sind – wie die anthropologischen und kunsttheoretischen Debatten des späten 19. und frühen 20. Jahrhunderts im Allgemeinen (vgl. Kappelhoff 2004) – maßgeblich von der Idee expressiver Bewegungsfigurationen beeinflusst. Ausgangspunkt ist die Annahme, Zeit und Bewegung würden im Sinne expressiver Gesten ausgestaltet – um darüber die Gefühle eines anonymen Publikums zu adressieren. *Die Expressivität der Bewegungsbilder selbst wird als ein polyphones Ganzes in Bewegung verstanden, dass sich auf Seiten der Zuschauenden direkt als ein Gefühl in der Zeit realisiert.*

- 2 Die im Folgenden vorgestellten methodischen Zugriffe führen zu einer Datengrundlage, die aus verschiedenen theoretischen Perspektiven explikatives Potenzial entfalten kann. In diesem Artikel soll es jedoch nicht darum gehen, die breite Nutzbarkeit der Daten aufzufächern, sondern den Zusammenhang zwischen theoretischen Prämissen, digitalen Methoden und analytischen Schlüssen entlang einer kohärenten Perspektive zu exemplifizieren.



1 Eisensteins Diagramm zu ALEXANDER NEWSKI

Der Ansatz, eine empirische Analyse audiovisueller Kompositionsmuster im Sinne der Ausdrucksbewegungstheorie vorzunehmen, ist jedoch zugleich mit großen methodologischen Herausforderungen verbunden. Besteht doch eine ihrer zentralen theoretischen Annahmen darin, die expressive Qualität von Bewegungen an die Idee temporaler Gestalten, d. h. die Idee einer Ganzheit der Bewegung in der Zeit, zu binden (vgl. Plessner 1982, 77). Aus einer solchen Perspektive ließe sich die expressive Qualität audiovisueller Bewegungsfigurationen allein durch die qualitative Deskription eines zusammenhängenden Ganzen fassen. Sind Ausdrucksbewegungstheorie und ein an empirische Daten audiovisueller Komposition gebundener filmanalytischer Ansatz somit unvereinbar?

Einen möglichen Ausweg aus diesem Dilemma weisen abermals die montage-theoretischen Überlegungen Eisensteins (2006a). Indem er Rhythmus und Bewegung konzeptionell verwebt, schlägt er eine Brücke zwischen den – als Rhythmen fassbaren – Mustern audiovisueller Komposition und gestalt-haften Zusammenhängen. Oder stärker methodologisch gewendet: Nimmt man eine Rekonstruktion der Rhythmik audiovisueller Sequenzen als Ausgangspunkt einer systematischen Deskription gestalthafter Ausdrucksbewegungen, so sollten sich Möglichkeiten eröffnen, die affektive Dynamik jener Sequenzen heuristisch fassbar werden zu lassen. Unter der Kategorie Rhyth-

mus fassen wir folglich ein temporales Muster, welches vom Zusammenspiel einer Vielzahl möglicher singulärer Akzente (Schnitte, Kamera- und Objektbewegungen, Sound-Effekte, musikalische Passagen usw.) geprägt wird.

Genau in diesem Sinne hat Eisenstein in seinen Ausführungen zur *Vertikalmontage* (Eisenstein 2006b) die grafische ‹Partitur› einer Filmsequenz angefertigt, welche sowohl einzelne Zeichen einer symbolisch begründeten Poetik als auch die – auditiven wie visuellen – Bewegungen, welche die ästhetische Erfahrung der Zuschauenden begründen, als Struktur zu fassen sucht (Abb. 1). Eine solche Partitur könnte auch im Kontext systematischer Methoden digitaler Filmanalyse den Ansatzpunkt für eine empirische Deskription audiovisueller Kompositionsmuster liefern. Diskrete und zeitlich definierte analytische Einzelbefunde auf unterschiedlichen Gestaltungsebenen werden als sich entfaltende Dynamiken präsentiert und der Blick wird auf deren Zusammenspiel gelenkt – lose vergleichbar mit der Partitur eines Orchesterstücks. So vermag es nicht zu überraschen, dass jüngere filmwissenschaftliche Forschungen den russischen Formalismus in prominenten Projekten Hand in Hand mit Methoden digitaler Filmanalyse erörtern und die Visualisierung von Analysedaten als Partitur wiederholt aufgreifen (vgl. Wurm 2009, 40 ff. sowie Heftberger/Loebenstein/Wasner 2009, 140).

Im Spannungsfeld von Tool und Methode

Vor diesem Hintergrund wollen wir nun aufzeigen, inwiefern methodologische Entwicklungen in der digitalen, computergestützten Filmanalyse einen Weg weisen können, jene kompositorischen Muster umfassend empirisch zu erfassen und vergleichend zu erforschen. Ansätze digitaler (und annotationsbasierter) Filmanalyse fokussieren meist auf einzelne oder wenige spezifische Gestaltungsebenen filmischer Expressivität wie z. B. die Schnittfrequenz (Tsivian 2009), Bewegung im Bild und Helligkeit (Cutting et al. 2011), Untertitel (Burghardt/Kao/Wolff 2016), Farbgestaltung (Burghardt/Kao/Walkowski 2018) oder Gesichtserkennung und Einstellungsgröße (Arnold/Tilton/Berke 2019). Aus unserer Perspektive besonders hervorzuheben sind die komplexen Anllysesystematiken und -umgebungen von Heftberger (2016) sowie Flückiger (2019), da sie am Zusammenspiel unterschiedlicher Gestaltungsebenen interessiert sind. Der im Folgenden ausgeführte Zugang der eMAEX-Systematik³ unterscheidet sich durch seinen

3 eMAEX steht für: *electronically based media analysis of expressive movement images*.

theoretischen Fokus auf die zeitliche Gestalt von Bewegungsbildern und seinen methodischen Zugang über ein maschinenlesbares, strukturiertes Analysevokabular.

Alle diese Ansätze bauen auf dem Versprechen auf, grundlegend neue Perspektiven zu eröffnen. Perspektiven, die von einem rein qualitativ-deskriptiven Vorgehen her, das ohne die komplexe Rekonstruktion audiovisueller Kompositionsmuster auf formalästhetischer Ebene auskommt, nicht ohne Weiteres zu erschließen sind. Die wichtigsten Motivationen sind zum einen, dass die empirische Dokumentation der kompositorischen Grundlagen audiovisueller Expressivität im transdisziplinären Austausch nutzbar gemacht werden kann, und zum anderen, dass systematische Vergleichsstudien und typologische Einordnungen in einer neuen Qualität durchführbar werden.

Dabei geht es auch darum, eine kritische Selbstreflexion aufrechtzuerhalten, die sich der Differenz einer solchen Datenbasis zur «Qualia» des geisteswissenschaftlich untersuchten Artefakts bewusst ist und sich gleichzeitig darauf bezieht (Drucker 2016, 246). D.h., es gilt die Datengenerierung und -visualisierung, wie Johanna Drucker fordert, als «Capta» (2011, 3), als Beobachter*innen-abhängigen Prozess zu denken und das Digitale nicht zum Selbstzweck werden zu lassen. Auch wenn es ein Aspekt der hier vorgestellten Herangehensweise ist, empirische Daten hervorzubringen, welche potenziell für *verschiedenste* Zugänge offen sind, so sollte auch deutlich werden, dass ohne ein *konkretes* theoretisches Modell zum Verhältnis von Film und Zuschauenden der Schritt von einem Set von Tools und systematischen Arbeitsroutinen hin zu einer *Methode* im vollen geisteswissenschaftlichen Sinne nicht zu vollziehen ist.

Im Folgenden möchten wir nun aufzeigen, wie die einzelnen Schritte einer, auf die hier skizzierten theoretischen Positionen aufsetzenden, filmanalytischen Methode aussehen können, welche Rolle digitale Tools für dieses Vorgehen spielen – und welche Einsichten sich darüber im Hinblick auf ein exemplarisches Beispiel ergeben.

Die digitale Rekonstruktion von Ausdrucksbewegung – eine methodologische Skizze

Wie also lassen sich das nachvollziehende Verstehen und die affektive Erfahrung von audiovisuellen Bildern aus den formalisierbaren Kompositionsmustern rekonstruieren? Nehmen wir als Beispiel ein von Sem Maltsev

2011 über die Plattform *Vimeo* veröffentlichtes eineinhalbminütiges Video aus dem Umfeld der *Occupy Wall Street*-Bewegung:⁴ Ein Spaziergang durch das Protestcamp am Zuccotti Park in Manhattan wird mit einer Montage von Wall-Street-Gebäuden und Statuen kontrastiert. Es bedarf keines großen Aufwandes, sowohl die von den Zuschauenden zu erschließende Ebene an Informationen und Bedeutungen – *Occupy* ist eine solidarische Basisbewegung, die gegen die Institutionen der Macht und des Finanzkapitalismus aufbegehrt – als auch die affektive Erfahrungsdimension einer schwebenden Erwartung, die in eine energetische, sich zuspitzende Kampfansage gegen «die da oben» umschlägt, zu benennen. Wie aber lassen sich die zugrundeliegenden Strukturen empirisch evident machen, sodass sich zum Beispiel die Bedeutungskonstruktionen und affektiven Qualitäten in der Folge systematisch mit anderen Untersuchungsgegenständen vergleichen lassen?

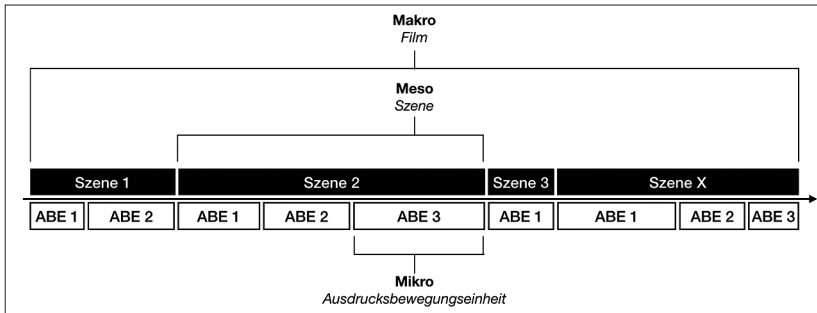
Um das Zusammenspiel von theoretisch geleiteten Arbeitsschritten und Tools bei der Qualifikation affektiver Dynamiken durch die Deskription audiovisueller Kompositionsmuster zu veranschaulichen, werden wir im Folgenden unsere methodische Grundlage – eMAEX – genauer vorstellen. Die methodologische und analytische Skizze soll das Zusammenspiel theoretischer Prämissen und digitaler Forschungsmethoden exemplarisch greifbar werden lassen.⁵ Dabei werden wir uns an zentralen Schritten des analytischen Zugriffs – der Segmentierung, Annotation, Visualisierung und Deskription audiovisueller Kompositionsmuster – orientieren.

Segmentierung

Eine der wesentlichen epistemologischen Fluchtlinien der Filmanalyse ist es, von der Zeitlichkeit des Gegenstands zu abstrahieren und einzelne Aspekte als poetische oder ästhetische Konzepte greifbar werden zu las-

4 Abrufbar unter <https://vimeo.com/30877096> (Letzter Zugriff am 03.02.2020). Die Kürze und relative Unterkomplexität des Beispiels sollen uns an dieser Stelle dazu dienen, die methodischen Schritte möglichst klar und konzis zu umreißen. Zugleich verbindet sich damit die Behauptung, dass auch umfangreichere und anspruchsvollere Formen audiovisueller Bilder in dieser Herangehensweise produktiv erschlossen werden können.

5 Der dieser Methode (eMAEX) zugrunde liegende Ansatz einer qualitativen Deskription audiovisueller Bewegungsfigurationen wurde bereits in früheren Publikationen dargelegt (Kappelhoff/Bakels 2011). Das hier dargelegte filmanalytische Vorgehen folgt der Weiterentwicklung von eMAEX im Rahmen der BMBF-eHumanities-Nachwuchsgruppe *Affektorhetoriken des Audiovisuellen* (siehe dazu auch: <http://www.ada.cinepoetics.fu-berlin.de>).



2 eMAEX Segmentierungsebenen

sen. Will man die Grundlage dieser Abstraktion auf Ebene der Zuschauerenerfahrung – die Wahrnehmung eines fortwährend in Veränderung befindlichen ästhetischen Objekts – empirisch adressieren, muss jene Zeitlichkeit strukturell fassbar werden.

Den Analysegegenstand systematisch zu segmentieren eröffnet eine derartige Möglichkeit und setzt gleichzeitig bereits ein Stück weit den Fokus jeder analytischen Untersuchung: Beschreibt man Filme als Abfolge von Einstellungen, entsteht ein anderer Gegenstand, als wenn man sich beispielsweise von akustischen Phänomenen leiten lässt. Indem in eMAEX die analysestrukturierenden Segmente als Erfahrungseinheiten (und damit als Synthese unterschiedlicher Gestaltungsebenen) verstanden werden, entsteht die Einteilung in Szenen als eine erste Ebene systematischer, überprüfbarer Abstraktion. Diese Segmente dienen letztendlich dazu, in ihrem zeitlichen Verlauf erfasst und anschließend im Hinblick auf ihre affektive Dynamik qualifiziert zu werden (s. u., Deskription). Die daraus resultierende Segmentierung als Ganzes erlaubt es dann, die Abfolge der szenischen Segmente von Filmen als affektdramaturgische Profile zu fassen (vgl. Kappelhoff/Bakels 2011, 88–90).

Die zweite Ebene unseres Vorgehens schließt direkt an die oben skizzierten theoretischen Prämissen an. Sie besteht darin, der Idee zeitlicher Gestaltwahrnehmung folgend, die im Film ausgemachten Szenen in sogenannte *Ausdrucksbewegungseinheiten* (ABEs) zu unterteilen. Demnach lassen sich Momente des gemeinsamen ›Auslaufens‹ oder ›Einsetzens‹ verschiedener rhythmisch-kinetischer Dynamiken des audiovisuellen Bildes, Momente der Kontrastierung oder ›Interpunktio‹ solcher Dynamiken auf Seiten der Zuschauenden als Grundlage einer Segmentierung filmischer Bilder in temporale Gestalten fassen (vgl. Bakels 2017, 155). Das dynamische Zusammenspiel der einzelnen

ABEs innerhalb einer Szene wiederum wird als *Meso-Ebene* gefasst und schafft einen konzeptionellen Übergang zwischen *Mikro-* und *Makro-Ebene*.⁶

Ausdrucksbewegungseinheiten werden dabei nicht a priori bestimmt, sondern werden systematisiert aus den kompositorischen Mustern der szenischen Gestaltung abgeleitet. Damit dies möglich wird, bedarf es allerdings eines Ansatzes der formalästhetischen Deskription, der die mannigfaltigen Ausdrucksebenen filmischer Bilder (wie z. B. Bildkomposition, Kameradynamik, Akustik oder Choreografie) umfasst, aus deren Zusammenspiel diese Muster greifbar werden. Hier kommen die Möglichkeiten digitaler Videoannotation ins Spiel.

Annotation

Will man die temporalen Gestalten nicht als ‹top-down›-Zuschreibungen, sondern vielmehr ‹bottom-up› fassen, gilt es zunächst, einen systematischen Zugriff im Sinne dessen, was in den *Computational Sciences* als *Data Modelling* konzeptualisiert wird, zu entwickeln. Notwendig ist eine Art und Weise der Datenerfassung, welche im Sinne einer konsistenten Datenqualität eindeutig ist und einen möglichst geringen und nachvollziehbaren Grad an Abstraktion und Interpretation aufweist. Mit Blick auf filmische Bilder bietet sich hier zunächst einmal die Formalästhetik audiovisueller Sequenzen an. Diese erweist sich bei genauerem Hinsehen als ein Amalgam technischer (wie z. B. im Falle von Bildformaten) und ästhetischer (wie z. B. bei der Benennung von Einstellungsgrößen) Begriffe, deren Definition zwingend eine Reihe von Entscheidungen und Setzungen für die tatsächliche Annotationsarbeit erfordert; gegenüber den komplexen hermeneutischen Operationen freitextlicher Filmanalyse bietet sich formalästhetisches Vokabular jedoch als vergleichsweise objektivierbare Terminologie an. Allerdings wirft ein derartiger Ansatz ein ganz eigenes Problem auf: jenes von Umfang, Komplexität und Skalierung.

So stand am Anfang unserer Bemühungen das Unterfangen, möglichst viele innerhalb der filmanalytischen Literatur adressierte formalästhetischen Dimensionen zu berücksichtigen. Mit sieben dieser Dimensionen – Bildkomposition, Kamera, Akustik, körperliche Expressivität, Montage, Sprache und Motive – erscheint die oberste Ebene unserer analytischen Routine noch überschaubar. Will man jedoch die einzelnen Binnendiffe-

6 Siehe Methodenteil in Müller/Kappelhoff 2018, 226–248 für eine detaillierte Beschreibung der Analyseebenen.

renzierungen dieser Auflistung erfassen, fächert sich diese grundlegende Ordnung innerhalb unserer Routine bereits in 80 Subdimensionen auf. Nimmt man schließlich noch hinzu, dass der Gedanke eines systematischen Data Modelling zwingend nahelegt, diese (Sub)Dimensionen im Sinne einer gleichbleibenden Terminologie (wie z. B. den einzelnen konkreten Benennungen von Einstellungsgrößen) zu klassifizieren, wird die Systematik abermals komplexer; im Falle unserer analytischen Routine ergeben sich darüber für die 80 Subdimensionen insgesamt über 400 definierte Werte.⁷

Eine derart komplexe analytische Routine für Spielfilme von teilweise über 120 Minuten Spieldauer potenziell *durchgehend* festzuhalten, sprengt die Kapazitäten etablierter filmanalytischer Hilfsmittel. Allerdings stehen mit einer Reihe von Software-Lösungen zur Videoannotation – wie etwa ELAN, VIAN oder ADVENE – mittlerweile geeignete digitale Tools zur Verfügung. Einem Videoschnittprogramm optisch nicht unähnlich, verfügen diese Programme meist über einen Videoplayer sowie über eine Ansichtsfläche, in der die mannigfaltigen (Sub)Dimensionen der Analyse als ‚Spuren‘ angezeigt werden; innerhalb dieser Spuren sind schließlich aus Sicht der Anwender*innen die jeweils annotierten Werte timecode-genau platziert.

Allerdings wirft die software-basierte Videoannotation auf einem Komplexitätsgrad wie diesem eine ganz eigene Problematik auf: Zwar lässt sich problemlos innerhalb des analytischen Rasters eines formalästhetischen Datenmodells ‹navigieren›, die Analyse ist jedoch weder in der Horizontalen (d. h. über den Zeitstrahl eines gesamten Films) noch der Vertikalen (d. h. über die bis zu 80 Spuren hinweg) als Ganzes ‹lesbar› (Abb. 3); entweder sprengt die Darstellung die Größe herkömmlicher Bildschirme, oder sie muss derart verkleinert werden, dass es unmöglich ist, noch einzelne Werte auszumachen.

Diesem Problem begegnen wir, indem wir unsere analytische Terminologie als eine *Ontologie*⁸ nach Maßgabe des Semantic Web aufsetzen (vgl. Hendl/Lassila/Berners-Lee 2001). D. h. auf der Ebene der digitalen Daten

7 Zur Erläuterung: Das Phänomen einer langsamen Kamerafahrt nach rechts würde hier aufgeschlüsselt als Zuordnung der Werte ‹Kamerafahrt›, ‹rechts› und ‹langsam› auf den Ebenen ‹Kamerabewegung Typ›, ‹Kamerabewegung Richtung› und ‹Kamerabewegung Tempo› zu dem entsprechenden Zeitabschnitt der Videodatei.

8 Die von uns – gemeinsam mit Henning Agt-Rickauer und Christian Hentschel vom Hasso-Plattner-Institut in Potsdam – entwickelte filmanalytische Ontologie lässt sich komplett und inklusive der Definitionen sämtlicher Ebenen und Werte unter <http://ada.filmontology.org> einsehen. Die AdA-Filmontologie steht unter der Creative Commons Attribution-ShareAlike 3.0 License, und kann frei genutzt und modifiziert werden.

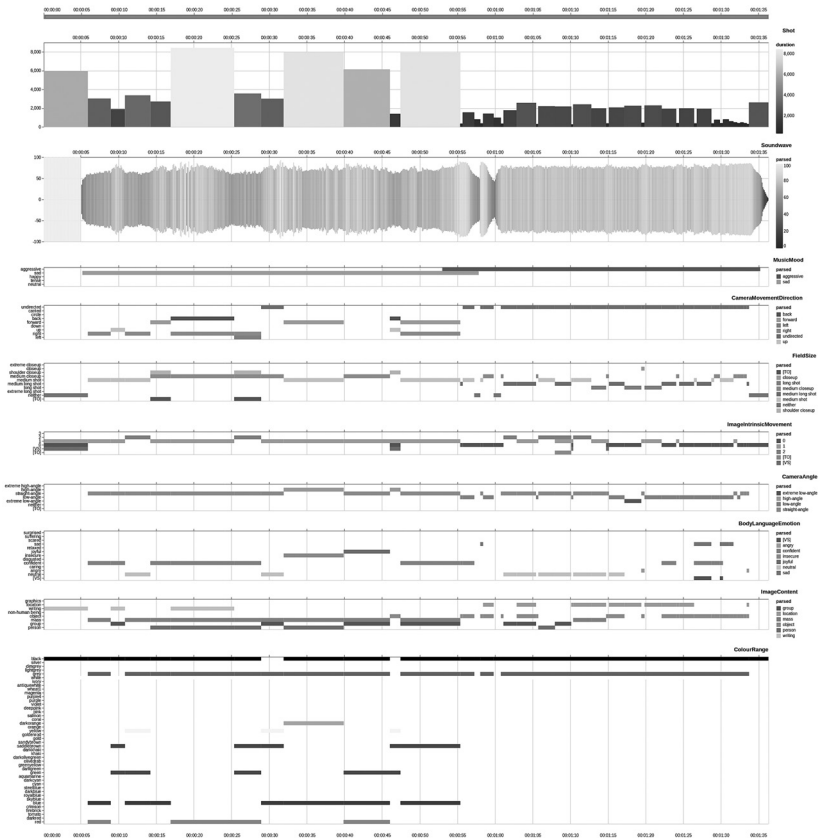
wird nicht mehr allein die timecode-basierte Annotation an sich markiert; vielmehr wird der jeweilige Wert als eine maschinenlesbare Relation der Annotation als Instanz der definierten Werte und Level gespeichert (vgl. Agt-Rickauer et al. 2018). Diese Erweiterung der vermerkten Daten – gewissermaßen hinter der Benutzeroberfläche – über das RDF-Datenformat (vgl. Decker/Mitra/Melnik 2000) macht es zum einen möglich, die Daten systematisch mit weiteren maschinenlesbaren Informationen (z. B. filmografischen Angaben) zu verknüpfen. Zum anderen sind damit die Relationen der verschiedenen Analyseebenen maschinenlesbar. Dies eröffnet die Möglichkeit, auf Entwicklungen im *Machine Learning* (vgl. Bishop 2006) zurückzugreifen. Sollte sich die Mustererkennung innerhalb einer 80 Ebenen umfassenden Systematik für die menschlichen Lesekapazitäten als zu komplex erweisen, so ermöglicht es uns das Datenformat perspektivisch, auf Algorithmen zurückzugreifen, um wiederkehrende Muster, Relationen usw. (semi)automatisch zu identifizieren. Zudem ermöglicht ein solches Data Modelling eine enge Verzahnung von manuellen und automatisierten Videoannotationen, indem automatisch generierte Erkenner-Ergebnisse auf der Grundlage eines definierten Begriffsgerüsts strukturiert werden können. Allerdings muss zum jetzigen Zeitpunkt betont werden, dass die bisher im Sinne von *Open Science* verfügbaren Erkenner bis auf wenige Ausnahmen (v. a. Schnitterkennung) aktuell noch nicht über die Präzision verfügen, welche für unsere Studien notwendig ist. Hier gilt es, Schnittstellen für die Zukunft und das systematische analytische Framework bereit zu halten, welches eine filmtheoretisch sinnvolle Einbindung von Erkennern ermöglicht.

Zugleich wollen wir die Frage audiovisueller Kompositionsmuster auf keinen Fall einfach in die Blackbox algorithmischen Lernens auslagern. Entsprechend haben wir den Weg beschritten, Visualisierungen unserer analytischen Datensätze in Erweiterung der verwendeten Videoannotationssoftware ADVENE⁹ zu entwickeln, die es uns ermöglichen sollen, kompositorische Muster zu vergleichen.

Visualisierung

Ansichtsoptionen gängiger Videoannotationssoftware eignen sich mit Blick auf die Lesbarkeit von komplexen Mustern audiovisueller Sequenzen nur bedingt. Dies ist vor allem darin begründet, dass die textorientierten

9 ADVENE ist eine frei verfügbare Open-Source-Video-Annotationssoftware (<https://www.advene.org>).



4 html-basierte Visualisierung der ADVENE-Annotationen zu OCCUPY WALL STREET

Ansichten dieser Programme primär für die Dateneingabe ausgelegt sind. Grafische Darstellungen (wie z. B. Bewegungsrichtungen als Pfeile) sind auf dieser Ebene nicht vorgesehen. Wie aber stellt man eine sinnvolle Integration von manuellen und automatisierten Eingabetools einerseits sowie Auswertung und Präsentation andererseits her? Die Analyseplattform VIAN, die unter der Leitung von Barbara Flückiger entwickelt wurde, stellt in dieser Hinsicht ein besonders gelungenes Beispiel für die umfassende Integration von Analyse- und Visualisierungstools in ein Ökosystem aus Annotationsprogramm, Onlineauswertung und Webpräsentation dar, wobei der Schwerpunkt auf den auf Farbanalyse spezialisierten Visualisierungen liegt (Flückiger 2017).

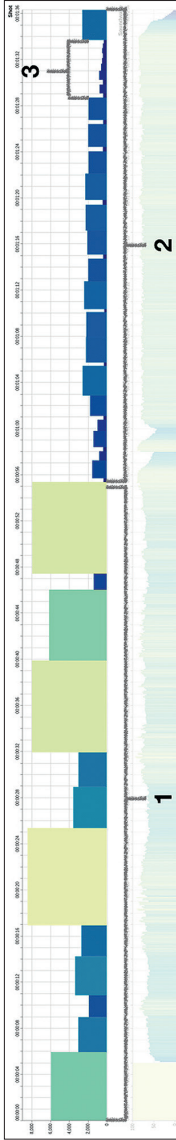
Für unser Vorgehen erwies sich die Verwendung des Semantic-Web-Datenstandards RDF als hilfreich. Da unsere Annotationswerte als Datum stets die Relation zu den jeweiligen (Sub)Dimensionen unserer analytischen Routine beinhalten, konnten wir – gemeinsam mit Olivier Aubert, einem der Entwickler von ADVENE – Visualisierungen entwickeln, welche sich als html-basierte Darstellungen in jedem gängigen Webbrowser anzeigen lassen (Abb. 4). Dabei lässt sich die Zoomstufe der Darstellung stufenlos anpassen, und es kann für jede der 80 Subdimensionen unserer analytischen Routine ebenso frei eine der verschiedenen Visualisierungsoptionen (Balkendiagramme, Histogramme, Kurvenverläufe, Symbole usw.) gewählt werden, wie sich die Auswahl und Reihenfolge der visualisierten Annotationsspuren beliebig verändern lässt. Auf diese Weise werden die Annotationen direkt als kompositorisches Muster lesbar. Für das hier gewählte kurze Beispiel ließe es sich folgendermaßen beschreiben:

Das eineinhalbminütige Video der *Occupy*-Bewegung bezieht zwei unterschiedliche Inszenierungsmodi aufeinander und erzeugt so eine Spannung: Das Histogramm in Abbildung 5 (bei dem Breite und Höhe der Blöcke der Länge der Einstellungen entsprechen) zeigt, dass die erste Hälfte mit deutlich längeren Einstellungen arbeitet (1) als die zweite Hälfte (2) und dass diese gegen Ende des Videos nochmals beschleunigt (3).

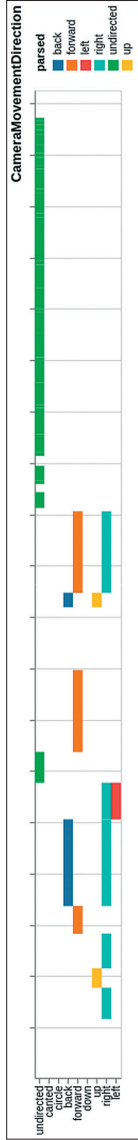
Die erste Hälfte mit den längeren Einstellungen wird von gleitenden, dynamischen Kamerabewegungen geprägt, während die zweite Hälfte stakka-tohaft nahezu statische Ansichten aneinanderreicht (auf der Ebene der Annotationen sieht man in Abbildung 6 eine Vielfalt an Bewegungsrichtungen in der ersten Hälfte gegenüber ungerichteten, kaum merklichen in der zweiten).

Auch auf akustischer Ebene ist ein Umschwung zu verzeichnen (siehe Abb. 7): Zwei Musikstücke wechseln sich in der Mitte des Spots ab. Die akzenthaften Paukenschläge des zweiten Stückes korrespondieren dabei auf visueller Ebene mit den rhythmisch wiederkehrenden kürzesten Einstellungen (gestrichelte Linien).

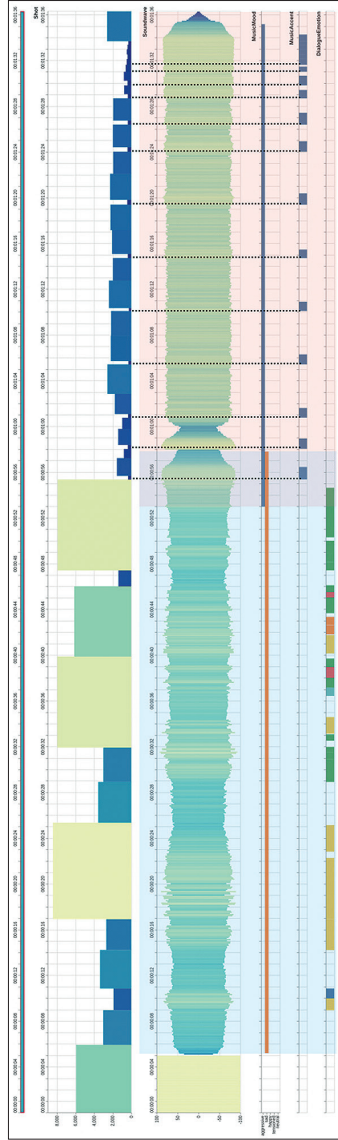
Die auf diese Weise gewonnenen Einsichten liefern eine empirische Grundlage, die den Gegenstand prägenden Bewegungsfigurationen systematisch als temporale Gestalten zu beschreiben und in ihrer affektiven Dimension zu qualifizieren. Statt einen einzelnen Parameter zu privilegieren, wird das multimodale Zusammenspiel beliebig vieler Gestaltungsebenen greifbar – man könnte für dieses Beispiel noch die Kamerawinkel oder die Mimik der abgebildeten Demonstranten und Statuen als instruktive Ebenen hinzufügen –, die einen zeitlich sich entfaltenden Prozess des Empfindens und des Verstehens strukturieren. Auf der Ebene einer Analyse von Spielfil-



5 Histogramm der Einstellungsängen



6 Kamerabewegungen



7 Zusammenspiel von Musikakzenten und Montage

men und längeren dokumentarischen oder essayistischen Formaten ergibt sich so die Möglichkeit, die Modulation der Zuschauendenwahrnehmung in der Zeit und über den Wechsel dominanter Parameter der Bewegungsfiguration hinweg in einer Darstellungsform greifbar zu machen. Zugleich ergeben sich völlig neue Perspektiven für die empirische Grundlage von vergleichenden Korpusanalysen, die motivische oder narrative Ankerpunkte mit auf Strukturen zielenden Gegenstandsbestimmungen vereinbar machen.

Deskription

Wie in den filmtheoretischen Erwägungen oben dargelegt, gehen die einschlägigen Theorien zur subjektiven Erfahrung expressiver Bewegungen von der Idee temporaler Gestalten aus. Während digitale Methoden die Analyse deutlich umfangreicherer Korpora ermöglichen und neue Perspektiven im Hinblick auf Typologien audiovisueller Figuretionen eröffnen, bedarf es demnach für den entscheidenden Schritt im Hinblick auf die affektive Qualität der jeweiligen Muster nach wie vor der qualitativ-deskriptiven Analyse durch Wissenschaftler*innen.

Mit Blick auf unser kurzes Beispiel lässt sich eine solche Deskription wie folgt formulieren: Über die Visualisierung war es uns möglich, das kompositorische Muster des *Occupy*-Clips in zwei ABEs einzuteilen: Die erste verbindet lange Einstellungen, langsame Kamerabewegungen mit sphärischen Chor- und Streicherflächen zu einem gleitend-schwebenden Bild harmonischen Beisammenseins; die zweite entfaltet im synchronen Zusammenspiel von schnellen Schnitten, Perkussions- und Bläserakzenten sowie einem zum Stakkato tendierenden, energetischen Streicherthema ein Bild des Getriebenseins, das von einem energetischen Rhythmus und dem Zulaufen auf eine Klimax getragen wird. Erst wenn wir das Muster in dieser qualitativen Deskription aufgreifen, offenbart sich im szenischen Zusammenspiel der Ausdrucksbewegungen das affektive ‚Tönen‘ der betrachteten Sequenz als die Dynamik einer Mobilisierung. Oder einfach gesagt: Computer fühlen nicht.

Perspektiven einer qualitativen Empirie audiovisueller Ästhetik

Die im Verlaufe dieses Artikels vorgestellten Überlegungen finden ihre gemeinsame Fluchtlinie in der Idee, wahrnehmungs- und affekttheoretischen Ansätzen der Filmtheorie ein Fundament zu verleihen, das sich über

formale Strukturen perspektiviert. Wie über die letzten Abschnitte bereits deutlich geworden sein sollte, geht es uns dabei jedoch keineswegs um eine ‹Empirie der Empirie halber›. Oder anders gesagt: Die hier dokumentierten methodologischen Anstrengungen finden ihre Motivation nicht im Unbehagen einer ‹weichen›, qualitativen Forschung gegenüber. Vielmehr lässt sich die Einsicht – Computer können Strukturen rekonstruieren, aber keine Gefühle –, mit der wir unsere methodologischen Überlegungen im vorherigen Abschnitt beendet haben, auch umkehren: Denn genauso gilt, dass sich letztlich nicht affektive Qualitäten audiovisueller Sequenzen vergleichend untersuchen lassen – vergleichen lassen sich nur die kompositorischen Muster, welche ihnen zugrunde liegen. Eine empirische Untersuchung audiovisueller Bilder als kommunikative Praxis bedarf somit unbedingt eines solchen Zugangs, der Strukturen nachvollziehbar macht. Sie bedarf aber selbstverständlich immer auch einer Einbettung in historische, institutionelle etc. Kontexte – Kritiken, Datenbanken zu Filmproduktion und -distribution etc. –, deren Verknüpfung mit den filmanalytischen Daten durch maschinenlesbare Datenformate via Linked Open Data ermöglicht wird.

Die qualitative Fluchtlinie dieses methodischen Vorgehens hat dabei nicht zuletzt auch den Zweck, die scheinbare Positivität der digitalen Wissensobjekte, die zwangsläufig in einem computergestützten Verfahren anfallen, zu relativieren und medienwissenschaftlich zu reflektieren. In diesem Sinne zielen die Verfahrensschritte und Werkzeuge nicht auf aggregierbare, statistisch auswertbare Daten, sondern darauf, dass sich datenbasierte Auswertung und hypothesengeleitetes, menschliches *pattern-seeking* gegenseitig informieren, eventuell sogar intervenieren, ohne einander zu determinieren. Dabei erwiesen sich die Feedbackschleifen aus Datenmodellierung und praktischer Annotationsarbeit sowie die Überprüfung der Intercoder-Reliabilität als wertvolle theoriebildende Prozesse, als elementarer Teil der Erkenntnisproduktion (vgl. Melgar Estrada et al. 2017, 62), schließlich zwingen sie uns dazu, die qualitativen Methoden zu präzisieren, die wir als ‹analoge› Praktiken nicht selten als selbstverständlich voraussetzen. Dass gegenwärtig die Grenzen automatischer Erkenner noch sehr deutlich sind, markiert in dieser Hinsicht nicht nur eine Herausforderung an die Aufwandskalkulation solcher Projekte, sondern auch eine Chance, die Daten konzeptuell von den Anforderungen der Filmwissenschaft und nicht von den Optionen der Algorithmen her zu denken.

Wir hoffen, dass wir in diesen knappen Ausführungen zeigen konnten, inwiefern es unter diesen Voraussetzungen fruchtbar und sinnvoll ist, die Untersuchung audiovisueller Bilder auf die Basis einer Auflösung von

Kompositionsmustern in einzelne strukturelle Ebenen und Wertebeziehungen zu stellen. Denn was an Filmen so schwer zu erklären ist, ist dass sie sich nie auf einzelne formalästhetische Eigenschaften reduzieren lassen, sondern immer nur mit Blick auf die zeitliche Entfaltung eines Ganzen Sinn ergeben.

Literatur

- Agt-Rickauer, Henning / Aubert, Olivier / Hentschel, Christian / Sack, Harald (2018) Authoring and Publishing Linked Open Film-Analytical Data. In: Proceedings of the 21st International Conference on Knowledge Engineering and Knowledge Management (EKAW 2018), Posters & Demo-Track.
- Arnold, Taylor / Tilton, Lauren / Berke, Annie (2019) Visual Style in Two Network Era Sitcoms. In: *Journal of Cultural Analytics*, S. 1–29, DOI: 10.22148/16.043.
- Bakels, Jan-Hendrik (2017) *Audiovisuelle Rhythmen. Filmmusik, Bewegungskomposition und die dynamische Affizierung des Zuschauers*. Berlin/Boston: De Gruyter.
- Balázs, Béla (2001) *Der sichtbare Mensch oder die Kultur des Films [1924]*. Frankfurt a. M.: Suhrkamp.
- Beilenhoff, Wolfgang (Hg.) (2005) *Poetika Kino. Theorie und Praxis des Films im russischen Formalismus*. Frankfurt a. M.: Suhrkamp.
- Bishop, Christopher M. (2006) *Pattern Recognition and Machine Learning*. New York u. a.: Springer.
- Bordwell, David (1989). A Case for Cognitivism. In: *Iris* 9, S. 11–40.
- Burghardt, Manuel / Kao, Michael / Walkowski, Niels-Oliver (2018) *Scalable MovieBarcodes – an exploratory interface for the analysis of movies*. IEEE VIS Workshop on Visualization for the Digital Humanities, Berlin.
- Burghardt, Manuel / Kao, Michael / Wolff, Christian (2016) Beyond Shot Lengths – Using Language Data and Color Information as Additional Parameters for Quantitative Movie Analysis. In: *Digital Humanities 2016: Conference Abstracts*. Jagiellonian University & Pedagogical University, Kraków.
- Cutting, James / Brunick, Kaitlin / DeLong, Jordan / Iricinschi, Catalina / Candan, Ayse (2011) Quicker, Faster, Darker: Changes in Hollywood Film over 75 Years. In: *I-Perception* 2,6, S. 569–576, DOI: 10.1068/i0441aap.

- Decker, Stefan / Mitra, Prasenjit / Melnik, Sergey (2000) Framework for the Semantic Web. An RDF tutorial. In: *IEEE Internet Computing* 4,6, S. 68–73.
- Deleuze, Gilles (1997) *Das Bewegungs-Bild. Kino 1* [frz. 1983]. Frankfurt a. M.: Suhrkamp.
- Drucker, Johanna (2011) Humanities Approaches to Graphical Display. In: *DHQ* 5,1, <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html> [letzter Zugriff am 17.02.2020].
- (2016) Graphical Approaches to the Digital Humanities. In: *A New Companion to Digital Humanities*. Hg. v. Susan Schreibman, Ray Siemens & John Unsworth. Chichester: John Wiley & Sons, S. 238–250.
- Eisenstein, Sergej M. (1988) *Das dynamische Quadrat. Schriften zum Film*. Hg. v. Oksana Bulgakova. Leipzig: Reclam, S. 10–16.
- (2006a) Die vierte Dimension im Film [1929]. In: Ders.: *Jenseits der Einstellung. Schriften zur Filmtheorie*. Hg. v. Felix Lenz. Frankfurt a. M.: Suhrkamp, S. 112–130.
- (2006b) Die Vertikalmontage [1940–1941]. In: Ders.: *Jenseits der Einstellung. Schriften zur Filmtheorie*. Hg. v. Felix Lenz. Frankfurt a. M.: Suhrkamp, S. 238–300.
- Flückiger, Barbara (2017) A Digital Humanities Approach to Film Colors. In: *The Moving Image* 17,2, S. 71–94.
- (2019) Avancierte Methoden der computer-gestützten ästhetischen Filmanalyse. In: *DHD 2019: Digital Humanities: multimedial multimodal*. Hg. v. Patrick Sahle. Frankfurt a. M.: Universitäten zu Mainz und Frankfurt, S. 12–21.
- Heftberger, Adelheid (2016) *Kollision der Kader: Dziga Vertovs Filme, die Visualisierung ihrer Strukturen und die Digital Humanities*. München: Edition Text und Kritik.
- / Loebenstein, Michael / Wasner, Georg (2009) Auf Spurensuche im Archiv. Ein Arbeitsbericht. In: *Digital Formalism. Die kalkulierten Bilder des Dziga Vertov*. Hg. v. Klemens Gruber, Barbara Wurm & Vera Kropf. (= *Maske und Kothurn* 55,3). Wien: Böhlau Verlag, S. 135–148.
- Hendler, James / Lassila, Ora / Berners-Lee, Tim (2001) The Semantic Web. In: *Scientific American*, 284,5, S. 34–43.
- Kappelhoff, Hermann (2004) *Matrix der Gefühle. Das Kino, das Melodrama und das Theater der Empfindsamkeit*. Berlin: Vorwerk 8.
- / Jan-Hendrik Bakels (2011) Das Zuschauergefühl. Möglichkeiten qualitativer Medienanalyse. In: *Zeitschrift für Medienwissenschaft* 5,2, S. 78–96.

- Kirsten, Guido (2016) Neoformalismus und Kognitive Filmtheorie. In: *Handbuch Filmtheorie*. Hg. v. Bernhard Groß & Thomas Morsch. Wiesbaden: Springer VS.
- Melgar Estrada, Liliana / Hielscher, Eva / Koolen, Marijn / Olesen, Christian Gosvig / Noordegraaf, Julia / Blom, Jaap (2017) Film Analysis as Annotation: Exploring Current Tools and their Affordances. In: *The Moving Image* 17,2, S. 40–70.
- Metz, Christian (1973) *Sprache und Film*. Frankfurt a. M.: Athenäum.
- Müller, Cornelia / Kappelhoff, Hermann (2018) *Cinematic Metaphor. Experience – Affectivity – Temporality*. Berlin/Boston: De Gruyter.
- Münsterberg, Hugo (1996) *Das Lichtspiel. Eine psychologische Studie (1916) und andere Schriften zum Kino*. Hg. v. Jörg Schweinitz. Wien: Synema.
- Plessner, Helmuth (1982) Die Deutung des mimischen Ausdrucks. Ein Beitrag zur Lehre vom Bewußtsein des anderen Ichs [1925]. In: Ders.: *Gesammelte Schriften VII*. Hg. v. Günter Dux. Frankfurt a. M.: Suhrkamp, S. 67–130.
- Sobchack, Vivian C. (1992) *The Address of the Eye. A Phenomenology of Film Experience*. Princeton NJ: Princeton University Press.
- Tsivian, Yuri (2009) Cinematics, Part of the Humanities' Cyberinfrastructure. In: *Digital Tools in Media Studies*. Hg. v. Michael Ross, Manfred Grauer & Bernd Freisleben. Bielefeld: Transcript, S. 93–100.
- Wurm, Barbara (2009) Vertov Digital. Numerisch-graphische Verfahren der formalen Analyse. In: *Digital Formalism. Die kalkulierten Bilder des Dziga Vertov*. Hg. v. Klemens Gruber, Barbara Wurm & Vera Kropf. (= *Maske und Kothurn* 55, 3). Wien: Böhlau, S. 15–44.