

# Reihe „Methodeninnovationen in der Kommunikationswissenschaft“

Empirische Methoden gehören zu den basalen Werkzeugen der Wissenschaft, lassen sich jedoch nicht wie Rezepte anwenden. Vielmehr sind im empirischen Forschungsprozess im Hinblick auf Fragestellung und Gegenstand zahlreiche spezifizierte Entscheidungen zu treffen. Mitunter müssen dabei neue, innovative oder nur selten genutzte Wege beschritten werden, etwa in Bezug auf das übergeordnete Untersuchungsdesign, die Stichprobengewinnung, Instrumenteneinführung, Datenerhebung oder die Auswertung der Daten und Befunde. Methodische Herausforderungen ergeben sich auch für Untersuchungsgegenstände z. B. im Rahmen computervermittelter Kommunikation. Mit diesen Aspekten sollen sich die Beiträge der von Wiebke Loosen konzipierten kontinuierlichen Reihe „Methodeninnovationen in der Kommunikationswissenschaft“ auseinandersetzen. Dabei kommen prinzipiell Beiträge aus allen Themenfeldern der Medien- und Kommunikationswissenschaft in Frage, die methodisch innovative Vorgehensweisen in eigenen Studien methodologisch reflektieren, andere Studien einer „sekundäranalytischen Methodenreflexion“ unterziehen und deren innovativen bzw. richtunggebenden Charakter herausarbeiten oder die unabhängig von konkreten Einzelstudien theoretisch-methodologisch Methodenentwicklungen nachzeichnen und reflektieren.\*

## Automatisierte Analyse medialer Bildinhalte

*Potenziale, Grenzen, methodisch-technischer Status Quo und zukünftige Herausforderungen – eine Bestandsaufnahme*

Stephanie Geise / Patrick Rössler / Simon Kruschinski

*Automatisierte inhaltsanalytische Verfahren haben bisher vor allem die Analyse von Medientexten bereichert. Zwar stellt die automatisierte Inhaltsanalyse von Bildern zum gegenwärtigen Zeitpunkt die wissenschaftliche Forschung vor größere Herausforderungen, als dass sie ihr Lösungen anbietet – doch konnten in den letzten Jahren (vor allem außerhalb der Kommunikationswissenschaft) fundamentale technische und methodische Weiterentwicklungen realisiert werden. Davon sollte auch die kommunikationswissenschaftliche Analyse von Medienbildern profitieren, weshalb dieser Beitrag darauf abzielt, eine Bestandsaufnahme der Methodik der automatisierten Analyse medialer Bildinhalte vorzunehmen, um die bisher weitgehend in anderen Disziplinen stattfindende theoretische und methodologische Diskussion der Kommunikationswissenschaft zugänglich zu*

---

\* Bislang sind in der Reihe Beiträge erschienen zu „Konstruktivismus und Methoden in der empirischen Sozialforschung“ (Verf. Armin Scholl, M&K 2/2011, S. 161–179), „Zur Verknüpfung manueller und automatischer Inhaltsanalyse durch maschinelles Lernen“ (Verf. Michael Scharzkow, M&K 4/2011, S. 545–562), zur „Visuelle[n] Inhaltsanalyse – ein Vorschlag zur theoretischen Dimensionierung der Erfassung von Bildinhalten“ (Verf. Stephanie Geise & Patrick Rössler, M&K 3/2012, S. 341–361), zur „Parasozialen Meinungsführerschaft als methodischer Herausforderung“ (Verf. Paula Stehr, Laura Leißner, Friederike Schönhardt & Patrick Rössler, M&K 3/2014, S. 395–416) und zur „Messung von Medienkompetenz – ein Testverfahren für die Dimension ‚Medienkritikfähigkeit‘ und die Zielgruppe ‚Jugendliche‘“ (Verf. Alexandra Sowka, Christoph Klimmt, Dorothée Hefner, Fenja Mergel & Daniel Possler, M&K 1/2015, S. 62–82).

*machen. Dabei reflektieren wir sowohl die Aufklärungspotenziale als auch die Grenzen des gegenwärtigen methodisch-technischen Status Quo und arbeiten auf dieser Basis zu künftige Herausforderungen für den Einsatz im Fach heraus.*

**Schlagwörter:** Automatisierung, Inhaltsanalyse, Bildinhaltsanalyse, automatisierte bzw. automatische Inhaltsanalyse von Bildern, computergestützte Verfahren, empirische Methoden der Kommunikationsforschung, visuelle Kommunikation, visuelle Kommunikationsforschung

## 1. Einleitung

Die Bedeutung von Medienbildern hat allein quantitativ enorm zugenommen: Minütlich werden knapp 60.000 Bilder auf Instagram hochgeladen, 140.000 Stunden Videomaterial auf YouTube betrachtet und drei Millionen Facebook-Einträge gepostet, von denen die Hälfte visuelle Elemente beinhalten (PennyStocks Labs 2015). Plattformen wie Facebook, Google, Snapchat, Flickr, Instagram oder Pinterest integrieren zunehmend auf Bildalgorithmen basierende Funktionen, etwa die bildbasierte Suche oder intelligente Gesichtserkennung (z. B. „DeepFace“ bei Facebook; Bohannon 2015). Für die Kommunikationswissenschaft ergeben sich aus diesen Entwicklungen neue methodische Herausforderungen bei der Datenerhebung und -auswertung: Wie lässt sich die Masse visueller Informationen inhaltsanalytisch erfassen? Welche Verfahren stehen dem For scher für die Automatisierung der Prozesse zur Verfügung? Kann der Computer überhaupt das „Sehen“ und Kategorisieren oftmals hochkomplexer und mehrdeutiger Bild inhalte „erlernen“?

Obwohl die Analyse von Medienbildern in der eher text-fokussierten Kommunikationswissenschaft lange Zeit eine untergeordnete Rolle gespielt hat, wurden in den letzten Jahren einige methodische Fortschritte in der Entwicklung der standardisierten Bildinhaltsanalyse vorgestellt (Geise & Rössler 2012, 2013a, 2013b; Grittmann & Lobinger 2011; Bock, Isermann & Knieper 2011; Bell 2010). Dass die Methodenentwicklung jedoch nicht als abgeschlossen gelten kann, zeigt auch die noch ausstehende Integration *automatisierter inhaltsanalytischer Verfahren*, die seit den Achtzigern (Klingemann 1984; Ballard & Brown 1982), besonders aber in den letzten Jahren zunehmend bei der Analyse *textlicher* Medienbotschaften eingesetzt wurden (vgl. Scharkow 2012, 2013; Sommer et al. 2014; Naab & Sehl 2014).

Eine Übertragung in die kommunikationswissenschaftliche Forschungspraxis gestaltet sich jedoch äußerst komplex, da sich der Fokus von Bildinhaltsanalysen außer auf formale Merkmale vor allem auf semantische Konzepte konzentriert, die für den Computer schwer zu „erkennen“ sind. So bietet die automatisierte Inhaltsanalyse von Bildern zum gegenwärtigen Zeitpunkt noch mehr Herausforderungen als Lösungen – und wird entsprechend als „holy grail of content-based media analysis“ (Chang 2002) bezeichnet. Doch signalisiert bereits die Zunahme an Publikationen den künftigen Stellenwert der automatisierten Bildinhaltsanalyse in der wissenschaftlichen Diskussion. Die Menge be willigter Forschungsprojekte zum Themenfeld „Mustererkennung“ des Bundesministeriums für Bildung und Forschung (BMBF) belegt daneben das rege Interesse der Wirtschaft an der „automatischen Erfassung, gezielten Erkennung und Verarbeitung von visuellen Medieninhalten“ (BMBF 2014). Vor allem *außerhalb* unseres Fachs wurden bereits fundamentale technische und methodische Weiterentwicklungen realisiert, von denen auch die Kommunikationswissenschaft profitieren kann, die die Inhaltsanalyse als ihre *originäre* Methode begreift und dementsprechend häufig einsetzt.

Vor diesem Hintergrund zielt der Beitrag auf eine Bestandsaufnahme der Methodik zur automatisierten Analyse medialer Bildinhalte, und zwar sowohl hinsichtlich deren Aufklärungspotenziale als auch der Grenzen des gegenwärtigen methodisch-technischen Status Quo. Dazu werden zunächst grundlegende Begriffe, Merkmale und Analyseverfahren entlang der technischen Analyselogik automatisierter Bildinhaltsanalysen eingeführt und dann in das kommunikationswissenschaftliche *Modell zur theoretischen Dimensionierung von Bildinhalten* (Geise & Rössler 2012) integriert, um zukünftige Herausforderungen für den Einsatz im Fach herauszuarbeiten.

## 2. Das Bild als Analyseobjekt der automatisierten Inhaltsanalyse

Die standardisierte Inhaltsanalyse dient der Untersuchung größerer Mengen an medialen Botschaften. Bei zunehmendem Materialkorporus wird die Anwendung von Verfahren automatisierter Inhaltsanalysen (AIA) attraktiv, denn sie reduziert einige Probleme der *quantitativen* Messung, insbesondere dass Daten mit Messfehlern behaftet und die basierenden Inferenzschlüsse mit stichprobenbedingter Unsicherheit verbunden sind (Scharkow 2011: 547). Die sequenziell-analytische Logik von Textkommunikation kommt der automatisierten Analyse dabei prinzipiell entgegen: Texte beruhen auf einem begrenzten, definitiven Satz von Zeichen (Buchstaben), die maschinell zuverlässig erkannt werden und deren Konstellationen (Worte) sich entsprechend des Forschungsinteresses in abgeschlossenen Listen (Diktiorär) für die automatisierte Vergleichsprüfung definieren lassen.

Bilder unterscheiden sich erheblich von numerischen oder lexikalischen Daten. Ein mit einer digitalen 4-Megapixel-Kamera geschossenes Urlaubsfoto kann genauso viele Daten enthalten wie zehn 400-seitige Bücher (Tönnies 2005: 18). Die fundamentale Differenz von Bildern ist aber, dass sie über eine holistisch-assoziative Präsentationslogik operieren, deren Struktur nicht über einfache „Regeln“ darstellbar ist. Dadurch besitzen Bilder wenige „manifest“ identifizierbare Merkmale; sie lassen sich auch nur bedingt in kleinere, hierarchisch-sequenzielle Einheiten zerteilen, wie dies etwa für Worte möglich ist (Müller & Geise 2015: 201; Geise & Rössler 2012). Da Bilder insofern aus *unstrukturierten Informationen* (Binary Large Objects) bestehen, sagt eine isolierte automatisierte Erfassung einzelner Bildelemente weder viel über die Bedeutung des erfassten Objekts noch über die Gesamtbedeutung des Bildes aus (Burger & Bürge 2005: 3). Auch die Erscheinungsform konkreter Phänomene im Bild erschwert die automatisierte Erfassung, da diese eine weitaus größere Varianz aufweisen. Bildinhalte und Bildaussagen sind damit sehr viel schwieriger zu kategorisieren als Textaussagen, deren begrenzter Zeichensatz sich durch vergleichsweise einfache Algorithmen abbilden lässt.

Ein Transfer automatisierter Verfahren offenbart hierbei eine erste Hürde: Zwar lassen sich *spezifizierte* visuelle Objekte bzw. Objektklassen heute relativ gut automatisiert erkennen, die große Variabilität vieler Objektkategorien bringt die Automatisierung aber an ihre Grenzen. Auch die Hintereinanderschaltung einzelner Objektidentifikations-Algorithmen bietet keine Lösung, da der Rechenaufwand pro Bild mit der Zahl der zu erkennenden Objekte enorm steigt. Bisher fehlt es also an Algorithmen, die unabhängig von einer spezifischen Objektklasse sind; meist werden daher vereinfachte Teillösungen realisiert (z. B. Fokus auf bestimmte typische Erscheinungsformen innerhalb der Objektklasse; Ommer 2013).

Bedeutsam für die Analyse von Bildern sind zudem ihre Bedeutungspotenziale, die in Relation zu situativen, individuellen, sozialen sowie medialen Kontexten stehen (Mitchell 1986) und ohne Berücksichtigung dieser oft nur unvollständig dechiffrierbar sind (Geise & Rössler 2012). Wie bei der AIA lässt sich hier durch Kombination manueller

und automatisierter Verfahren zusätzliche Analysetiefe gewinnen (Wettstein 2014): Das Vorwissen des Forschers/Codierers, insbesondere seine ikonographisch-ikonologische Verknüpfung mit anderen Bildmotiven, können nicht nur bei der Interpretation und Einordnung helfen, sondern auch in neuronale Netze künstlicher Intelligenz eingespeist werden (vgl. Zeiler & Fergus 2014). Hier steht die Forschung zwar noch am Anfang, doch stellt die computergestützte Indexierung von Inhalten, der Abgleich ähnlicher Inhalte innerhalb von Bildern oder zwischen verschiedenen Bildern bereits heute ein wichtiges Instrument der automatisierten Inhaltsanalyse von Bildinhalten dar (*content-based image retrieval* oder *content-based visual information retrieval*; Smeulders et al. 2000; Datta, Li & Wang 2005). Die Verknüpfung von Einzelbild und breiterem Kontext hat sich bei der Analyse von großen Datenmengen – etwa bei der Analyse von Zeitungsartikeln mit Bildmaterial oder von Schlüsselereignissen (Rauchenzauner 2008) – vielfach als unverzichtbar erwiesen.

### 3. Eine Bestandsaufnahme zur automatisierten Bildinhaltsanalyse (ABIA)

Auseinandersetzungen mit der automatisierten Analyse von Bildinhalten lassen sich vor allem in der Informatik finden, speziell in den Bereichen *Image Processing* (Bildverarbeitung), *Pattern Recognition* (Mustererkennung) und *Computer Vision* (Maschinelles Sehen). Jeder Fachbereich arbeitet mit verschiedenen algorithmischen Verfahren, doch werden diese auch übergreifend eingesetzt. Das grundlegende Prinzip aller Verfahren ist dabei, Bildinformationen durch computergestützte Berechnungen 1) zu extrahieren, 2) zu annotieren und 3) in unterschiedlichen Graden zu „interpretieren“ (Tönnies 2005: 15). Die ABIA bezeichnet insofern computergestützte Verfahren, die darauf abzielen, digitalisierte visuelle (d. h. abgebildete) Information mit einem zuvor definierten Modell der realen Welt in Bezug zu setzen (Sonka, Hlavac & Boyle 2014: 5). Die Bildinformationen werden dazu auf verschiedenen Ebenen – startend bei der rohen Bildinformation, hin zu logisch-determinierten Gruppierungen und Kontextrelationen – analysiert und die verarbeiteten Daten mit einem internen Modell abgeglichen. Dabei werden komplexe semantische Konzepte in computerlesbare Algorithmen überführt, um die Inhalte von Multimediaobjekten zu „erkennen“ bzw. ihnen automatisiert Bedeutungen zuzuschreiben. Von maschinellem „Bildverständnis“ wird gesprochen, wenn das der Maschine inhärente Modell der Realität entspricht (ebd.: 465); streng genommen liegt hierbei aber lediglich eine *erfolgreiche visuelle Mustererkennung* und *Musterzuordnung* vor. Für die Klassifizierung visueller Muster werden die Bildobjekte üblicherweise zunächst in typische Referenzfälle eingeteilt und zu Referenzklassen gebündelt, die in das System eingespeist werden. Auf dieser Basis erfolgt ein visueller Mustervergleich zwischen den abgebildeten Bildobjekten und den gespeicherten Referenzfällen.

Die Auswahl des Analyseverfahrens und der entsprechenden Algorithmen trifft der Computer anhand von *Features*. Features sind numerische oder qualitative Merkmale des materiellen Analyseobjektes, die aus Bildpunkten (Werten und Koordinaten) berechnet und mit geeigneten Algorithmen extrahiert werden (Tönnies 2005). Features können 1) technisch-operational in *Low- und High-Level-Features* und 2) strukturell in *Lokal- und Global-Features* eingeteilt werden. Die erste Unterscheidung betrifft den semantischen Gehalt: Während Low-Level-Features primär formale Aspekte des Bildinhalts (Kanten, Texturen) oder statistische Messgrößen (Farbwert, Färbung) beschreiben, beziehen sich High-Level-Features auf zusammenhängende-interpretative Aspekte des Bildes. Lokal-Features beschreiben einzelne Bildregionen und können im Bildobjekt oder an dessen Außenkontur (Linienstücke, Kreisbögen oder Eckpunkte) lokalisiert werden. Globale Features adressieren hingegen bestimmte Charakteristika (Histo-

gramm-Merkmale, spektrale Merkmale, geometrische Merkmale) des gesamten Bildes (Tönnies 2005: 296ff.).

Automatisierte Verfahren und Algorithmen der Bildinhaltsanalyse wurden v. a. in den MINT-Disziplinen kontinuierlich weiterentwickelt. In der Medizintechnik wurden etwa Verfahren vorgestellt, Mikroskopie-, Kernspin- und Röntgenbilder mit maschinelner Unterstützung auszuwerten (z. B. Sharma & Aggarwal 2010). Auch der zunehmende Präventionsbedarf in der inneren und äußereren Sicherheit, die zunehmende bildliche Datenmenge online sowie preisgünstige Massendatenspeicher motivierten verstärkte Forschungsaktivitäten. So fördert das BMBF im Themenfeld „Mustererkennung“ seit 2008 elf Verbundprojekte, die Innovationen zur automatisierten Erfassung, Erkennung und Verarbeitung von Bilddaten aus unterschiedlichen Quellen (z. B. Kameras, Sensoren, Scanner) generieren.

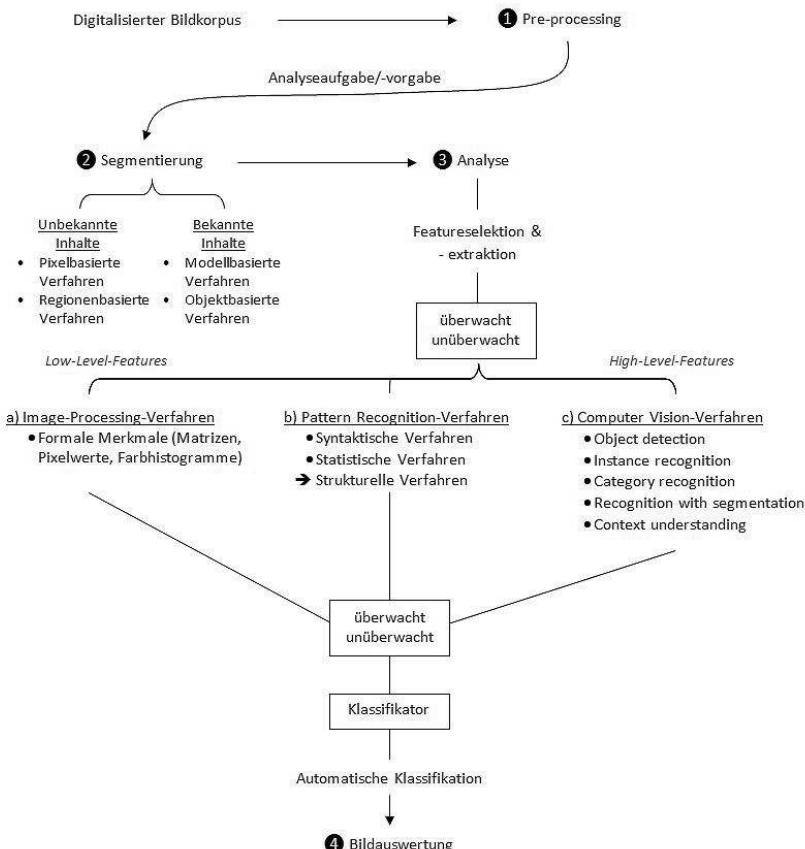
Auch im Kontext kommunikationswissenschaftlicher Fragestellungen wurden automatisierte Bildinhaltsanalyseverfahren wiederholt reflektiert, bisher aber nur selten eingesetzt. Effelsberg (1999) diskutierte Einsatzmöglichkeiten von Computeralgorithmen, die Basisparameter wie Farbverläufe, Kanten- oder Gesichtserkennung aus visuellen Daten extrahieren, schlussfolgerte jedoch, dass „ein tiefes [inhaltliches] Verstehen [...] durch den Computer auf absehbare Zeit nicht möglich sei“ (ebd.: 181). Ein Forschungsfeld, in dem wiederholt automatisierte Verfahren zum Einsatz kamen, ist die Analyse von *Visual Bias* – etwa hinsichtlich der Frage, welche Politiker wie oft in den untersuchten Nachrichtensendungen auftraten (Berg et al. 2004; Ozkan & Duygulu 2006). Auch die automatisierte Identifikation von pornografischen Bildinhalten (Lefebvre et al. 2006; Zheng et al. 2004) oder von filmanalytischer Stilistik in (Nachrichten)-Videos (Bastan & Duygulu 2006; Salt 2006) integrierte Fragestellungen der empirischen Kommunikationsforschung.

Vor allem jüngere Studien geben durch die Verwendung von selbstlernenden Algorithmen, die sich die Semantik von Bildinhalten in einem iterativen Verfahren aneignen, einen Vorgeschmack auf die zukünftigen Anwendungsmöglichkeiten für die Kommunikationswissenschaft. So implementierten Joo, Steen und Zhu (2015) einen Algorithmus, der US-Politikern automatisiert Charaktereigenschaften anhand ihrer Portraits zuwies und davon ausgehend auf den Wahlausgang schloss. Joo, Li, Steen und Zhu (2014) stellten einen Algorithmus vor, der Politikerbilder in der Nachrichtenberichterstattung auf ihre „Vorteilhaftigkeit“ und persuasive Wirkung interpretiert (positives vs. negatives Politikerbild). Portraits und typische Politiksettings und -gesten, die als Analyseobjekte dieser Forschungsarbeiten dienten, bieten sich allerdings aufgrund der sehr gut entwickelten Gesichtserkennungsalgorithmen und der begrenzten Variabilität des Form- bzw. Symbolrepertoires besonders gut für automatisierte Analysen an.

#### *Dimensionierung von Verfahren der automatisierten Inhaltsanalyse von Bildinhalten entlang ihrer Analyselogik*

Die meisten automatisierten Verfahren folgen einem Analyseprozess, der auf einer vierstufigen Analyselogik nach Ballard und Brown (1982) basiert (Abb. 1). Die in den Rohdaten enthaltenen Informationen werden mit steigender Analysetiefe graduell verdichtet, wobei die Komplexität verwendeter Algorithmen zunimmt. Da diese Idee anschlussfähig an zentrale kommunikationswissenschaftliche Arbeiten zum Thema erscheint (vgl. Scharkow 2012; 2013; Keyling 2014; Günther & Quandt 2015), ziehen wir sie im Folgenden heran, um den Beitragsgegenstand zu dimensionieren, Verfahren der automatisierten Bildinhaltsanalyse zu verorten und schließlich ihr kommunikationswissenschaftliches Anwendungspotenzial zu prüfen.

Abbildung 1: Analyselogik und Analyseverfahren der automatisierten Inhaltsanalyse von Bildinhalten



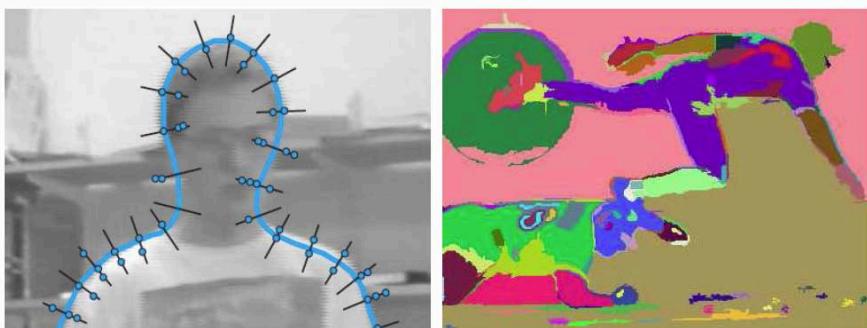
Quelle: eigene Darstellung.

(1) Beim *pre-processing* werden aus dem digitalisierten Materialkorporus Bildinformatio-nen als Rohdaten erfasst (iconic image) und für den nachfolgenden Analyseprozess op-timiert; durch die Bildaufnahme verursachte Bildstörungen (z. B. Über-/Unterbelich-tung, Unschärfe, Kontrast, Bildrauschen) werden beseitigt, bestimmte Bildbereiche (Kanten, Regionen) hervorgehoben. Hierbei wird auf Grundlage einzelner Pixel gear-beitet, die isoliert zu wenige Informationen beinhalten, um Aussagen über die im Bild befin-dlichen Objekte zu treffen (Tönnies 2005: 198).

(2) Bei der *Segmentierung* werden die Rohdaten in semantische Einheiten innerhalb von Gruppen segmentiert, die aufgrund ihrer Eigenschaften (z. B. Größe, Farbe, Form, Kanten, Regionen, geometrische Repräsentation) mit einer gewissen Wahrscheinlichkeit zum gleichen abgebildeten Objekt gehören. Die formale Ebene der syntaktischen In-formation wird hierbei mit der symbolischen Ebene der semantischen Information ver-bunden, auf der Erklärungen für das Bild oder einzelne Bildelemente gesucht werden.

Da derartige *pixel- bzw. regionenbasierte* Segmentierungsverfahren (Abb. 2) kein Wissen über die *inhaltliche* Bedeutung der Bildelemente voraussetzen, sind sie auf nahezu jedes Bild anwendbar. Allerdings sind sie weit von einer „Interpretation“ der Bildinhalte entfernt, weshalb sie für komplexe Abbildungen bzw. Fragestellungen nur bedingt geeignet sind (Tönnies 2005: 199). Dennoch kann auch bedeutungsrelevantes Wissen über die erwarteten Objekte in die Segmentierung eingebunden werden; *modell- oder objektbasierte Segmentierungen* (Abb. 3) arbeiten hier mit Suchstrategien, die mithilfe von „Vorwissen“ über ein bestimmtes Objekt Modelle mit den zu untersuchenden Bilddaten abgleichen, um bestimmte Bildinhalte zu „erkennen“.

Abbildung 2: *Pixel- und regionenbasierte Segmentierungsverfahren*



Quelle: Isard & Blake 1998: 15. © 1998 Springer.

Abbildung 3: *Modell- und objektbasierte Segmentierungsverfahren*



Quelle: Boykov & Funka-Lea 2006: 124. © 2006 Springer.

(3) Bei der eigentlichen *Analyse* werden die segmentierten Bildobjekte über ihre Bildmerkmale zueinander in Bezug gesetzt (Sonka, Hlavac & Boyle 2014: 101). Für jedes Segment werden dazu Features identifiziert, durch die „ein bestimmter Aspekt der Objektbedeutung beschrieben wird“ (Tönnies 2005: 296). Als Signatur einer Bildregion wird zudem für jedes Bildobjekt ein *feature vector* berechnet, der diese möglichst eindeutig und trennscharf beschreibt (Burger & Bürge 2005: 218-219). Durch Algorithmen,

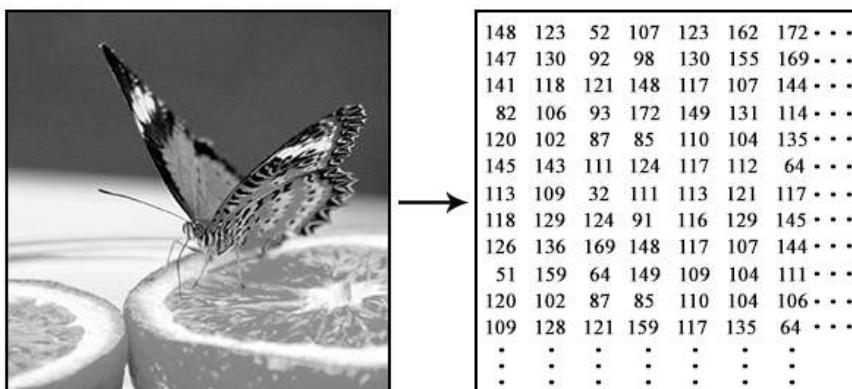
die Merkmale anhand ihrer Kanten, Ecken, Farbe oder Textur auf lokaler oder globaler Ebene identifizieren, werden Bildpunkte über ihre jeweiligen Ausprägungen gruppiert.

(4) Bei der *Bildauswertung* werden aus den gebildeten Einheiten Informationen extrahiert und auf Grundlage eines zuvor trainierten Klassifikators in Ähnlichkeitsklassen eingeteilt. Die Objektklassifikation beruht also auf einer automatisierten Zuordnung, über die den einzelnen Objekten bzw. Segmenten eine möglichst optimale Erklärung zugeordnet wird. Auch bei dieser Form des maschinellen Lernens lassen sich unüberwachte und überwachte Klassifikationsverfahren unterscheiden (Hüftle 2006; Pernkopf 2004). Beim überwachten Lernen werden aus der Datenbasis Objekte zufällig ausgewählt und zu einer Trainingsmenge zusammengestellt. Jedes Trainingsobjekt wird einer vorher festgelegten Klasse zugewiesen; von den vorklassifizierten Fällen „lernet“ der Algorithmus, zu welcher Klasse noch unbekannte Objekte gehören (Pernkopf 2004: 8). Unüberwachte Klassifikationsverfahren suchen stattdessen nach „natürlichen Gruppen“ (ebd.: 30), wobei Objekte auf Basis von Dichte- und Wahrscheinlichkeitsberechnungen klassifiziert werden. Die Daten werden als Modellinformationen gespeichert und können mit späteren Klassifikationen unbekannter Bilder verglichen werden.

Zusammengefasst liegen also bereits einige Lösungen vor, um klar definierbare Inhalte von Stand- und Bewegtbildern automatisiert zu erfassen:

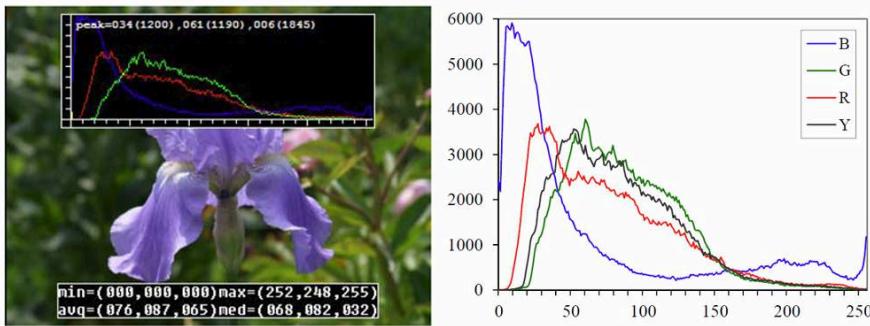
(a) *Image-Processing*-Verfahren, die der Hervorhebung, Identifikation und Extraktion digitaler Bildinformationen (meist ausschließlich) auf Basis von Pixeldaten dienen (Tönnies 2005: 26), werden erfolgreich in verschiedenen Ingenieursdisziplinen eingesetzt (Burger & Bürge 2005: 3). Ihr Fokus liegt allerdings auf der Extraktion von Low-Level-Features (siehe Abb. 4, Abb. 5); semantische Bedeutungen können dabei nicht extrahiert werden. Der *semantic gap* – „the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation“ (Smeulders et al. 2000: 5; vgl. Stommel & Müller 2011: 246) – ist folglich groß.

Abbildung 4: Anwendungsbeispiel *Image-Processing*: Übergang von einem Ausgangsbild zu einer zweidimensionalen Matrix



Quelle: eigene Darstellung.

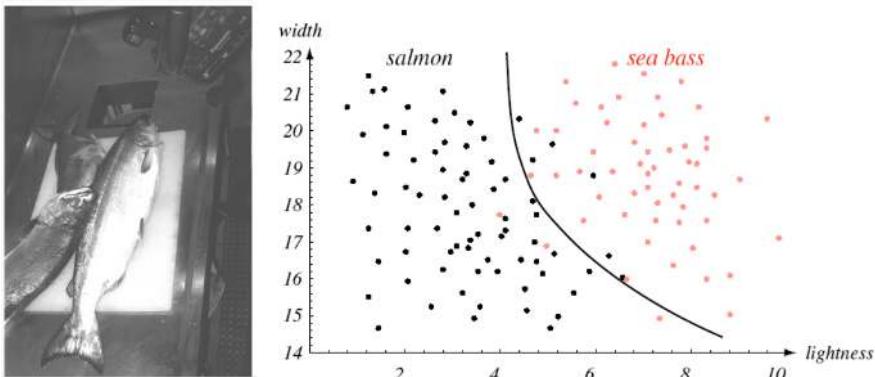
Abbildung 5: Anwendungsbeispiel Image-Processing: Histogrammanalyse eines Bildes mit entsprechendem Belichtungshistogramm



Quelle: Szeliski 2010: 108.

(b) Verfahren der *Pattern Recognition* gehen hier einen Schritt weiter und versuchen, in Bildern Muster, d. h. Regelmäßigkeiten, Wiederholungen, Ähnlichkeiten oder Gesetzmäßigkeiten zu identifizieren. Mit Franz (2007) sind drei Ansätze zu differenzieren: 1) *Syntaktische* Verfahren „erkennen“ Muster auf Grundlage von algorithmisch berechneter Ähnlichkeitsregeln, um Objekten der gleichen Kategorie dieselben Beschreibungen zuzuweisen. 2) Bei *statistischen* Verfahren erfolgt die Zuteilung von Mustern durch errechnete Wahrscheinlichkeitsklassen, um zu entscheiden, ob ein Objekt zu der einen oder anderen Kategorie gehört (Abb. 6). 3) Die *strukturelle* Mustererkennung kombiniert Verfahren beider Ansätze, führt die jeweiligen Einzelergebnisse zusammen und berechnet daraus die Klassenzugehörigkeit. So erkennen *Bayes'sche* Netze zuverlässig einzelne Gesichtsteile und errechnen daraus die Objektklasse „Gesicht“ (ebd.: 17).

Abbildung 6: Anwendungsbeispiel Statistisches Pattern-Recognition-Verfahren und Objektkategorie: Lachs oder Seebarsch



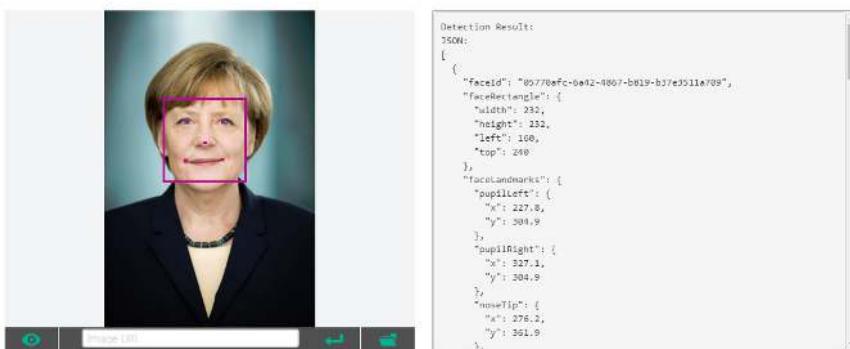
Quelle: Franz 2007.

(c) Hierauf aufbauend beschäftigt sich die *Computer Vision* schließlich mit dem Problem, Informationen über Objekte der realen dreidimensionalen Welt aus zweidimensionalen

Bildern zu gewinnen (Sarfraz 2014). Ziel ist, Prozesse der menschlichen visuellen Wahrnehmung und Verarbeitung (Human Vision) möglichst maschinell ab- bzw. nachzubilden (Marr 1983; Tönnies 2005), sodass der Computer „lernt“, abgebildete Objekte *ähnlich* wie das menschliche Gehirn zu erkennen und zu interpretieren (Kohle 2013: 72f.). Während Mustererkennungsverfahren häufig aufgrund der geringen und oft widersprüchlichen Informationen über den Inhalt des Gesamtbildes an Grenzen stoßen, versuchen Computer-Vision-Algorithmen die kompositionelle Regelhaftigkeit der Welt und des menschlichen Sehens zu nutzen. Die Grundidee ist hierbei, dass fast alle visuellen Objekte aus „comparably few, simple, and widely usable parts“ (Ommer & Buhmann 2007: 1) bestehen – etwa aus Punkten, Strichen oder Kreisen sowie ihren Beziehungen zueinander. Ein Computer-Vision-Algorithmus verarbeitet die Lage der einzelnen Bildobjekte zueinander und gruppiert Bildbereiche, die zum selben Bildobjekt passen, während er zugleich eine passende „Interpretation“ für das gesamte Objekt sucht. Derartige Algorithmen können bereits anhand eines kleinen Satzes von Trainingsbildern lernen, die vielen Bildpunkte eines Bildes zu bedeutungstragenden Kompositionen zu gruppieren; gleichzeitig verbessert der Computer seine Fähigkeit, Charakteristiken zu erkennen, mit denen er eine Objektkategorie von einer anderen unterscheiden kann (Ommer 2011; Bell et al. 2013). Dabei stellen *object detection*, *instance recognition*, *category recognition*, *recognition with segmentation* und *context understanding* entscheidende Ansätze dar, um Abbildungsgegenstände wie Personen, Objekte und Situationen automatisiert zu erfassen und zu identifizieren.

Bei der *object detection* wird auf Grundlage von Image-Processing- oder Pattern-Recognition-Verfahren entschieden, ob in einem analysierten Bild ein bestimmtes Objekt (z. B. ein Gesicht) abgebildet ist (Abb. 7). Über Gesichtsdetektionsalgorithmen, die auf Verfahren der strukturellen Pattern Recognition basieren, können im Materialkorporus beispielsweise bestimmte Personen automatisiert erfasst werden (Szeliski 2010); so lassen sich z. B. bekannte Politiker in der medialen Bildberichterstattung automatisiert inhaltsanalytisch erfassen (Joo, Steen & Zhu 2015).

Abbildung 7: Anwendungsbeispiel Object Detection: Gesichtserkennung mit Microsofts ProjectOxford-Algorithmus



Quelle: ProjectOxford 2015.

Abbildung 8: Anwendungsbeispiel Instance Recognition: Automatisiertes Wiederfinden von Objekten in einem Bild



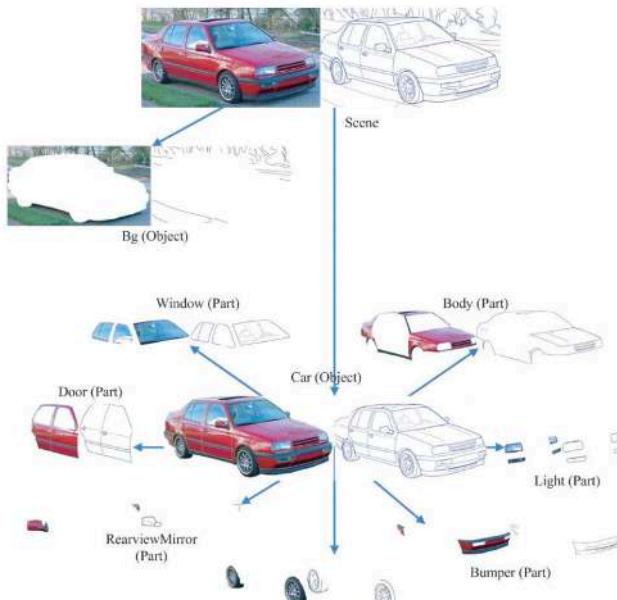
Quelle: Lowe 2004: 23. © 2004 Springer.

*Instance Recognition*-Ansätze greifen auf Bilddatenbanken zu und vergleichen auf Basis von Low-Level-Features, ob hinterlegte Objekte im Bild vorhanden sind (Abb. 8). Um den Abgleich zu erleichtern, sind die einzelnen Bildobjekte in der Datenbank meist in unterschiedlichen Perspektiven und vor einem gleichfarbigen Hintergrund gespeichert. Eine Objektidentifikation erfolgt, wenn eine ausreichend große Menge an Features im Bild einem bestimmten Datenbankobjekt zugeordnet werden kann. In der Praxis stößt dieses Verfahren allerdings noch in Fragen der Durchführungsdauer an Grenzen, wenn die in der Datenbank abgespeicherten Objekte in die Millionen gehen (ebd.: 687).

Im Gegensatz zu den bisher vorgestellten Verfahren erlauben *Category Recognition*-Ansätze auch tiefergehende Bildinhaltsanalysen, da sie nicht nur die Präsenz/Absenz eines Objektes in einem Bild untersuchen, sondern einzelne Bilder „in Bezug auf eine größere Anzahl von Bildkategorien analysieren, die entweder einzelnen Objekten oder verschiedenen Typen von Szenen entsprechen“ (Stommel & Müller 2011: 258). Szeliski (2010: 658) beschreibt diese Ansätze als äußerst anspruchsvoll, weil hier Computeralgorithmen in einem Bild extrem unterschiedliche Klassen – vom Hund über Möbel bis hin zu Autos – erkennen müssen. Auch *Category Recognition*-Ansätze benutzen Bilddatenbanken als Grundlage für das Erlernen und Erkennen von Features (eine Übersicht findet sich in Szeliski 2010: 719ff.). Zwei vielversprechende *Category Recognition*-Ansätze sind das *bag of words*-Verfahren, das eine Übertragung von statistischen Wörterbuchbasierten Textanalyseverfahren auf visuelles Material darstellt (ebd.: 697ff.), und die *part-based recognition*, die auch bei der Gesichtserkennung zum Einsatz kommt (Abb. 9). Beide Ansätze identifizieren individuelle Features (z. B. Augen, Nase, Mund) sowie ihre räumliche Lage zueinander und schließen dadurch auf deren übergeordnete Bedeutung (z. B. Gesicht).

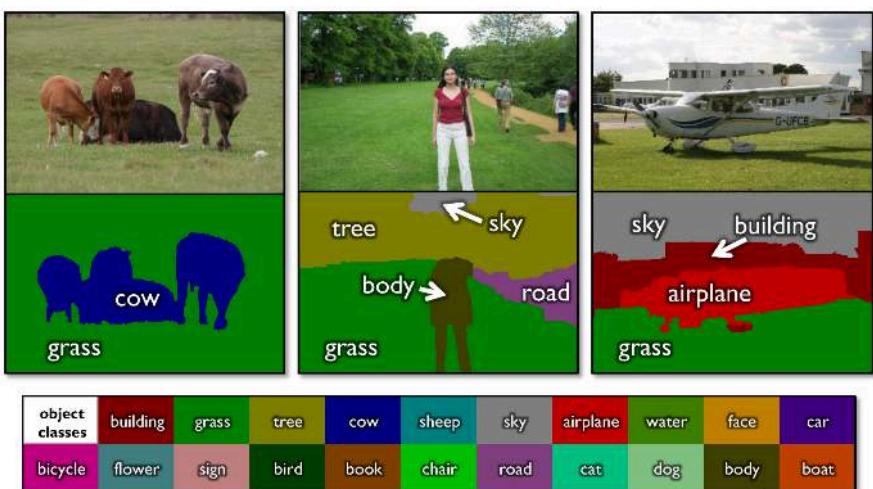
Als „most challenging version of generic object recognition“ bezeichnet Szeliski (ebd.: 704) jedoch die *recognition with segmentation*, die zur Erkennung von Objekten automatisierte Segmentierungen durchführt (Abb. 10). Dafür werden zu untersuchende Bilder in kleinere Teile segmentiert, die mit „pre-learned (segmented) object models“ (ebd.: 706) aus einem Trainingsdatensatz verglichen werden (vgl. Ferrari et al. 2006). Vor allem durch ihre verschiedenen Abwandlungsformen, die eine Segmentierung vor oder zeitgleich zur Klassifizierung erlauben, besitzen derartige Verfahren ein sehr großes Potenzial für die automatisierte Detektion von Bildobjekten.

Abbildung 9: Anwendungsbeispiel Category Recognition: Part-based Recognition Modell eines Autos



Quelle: Zhu & Mumford 2006: 275.

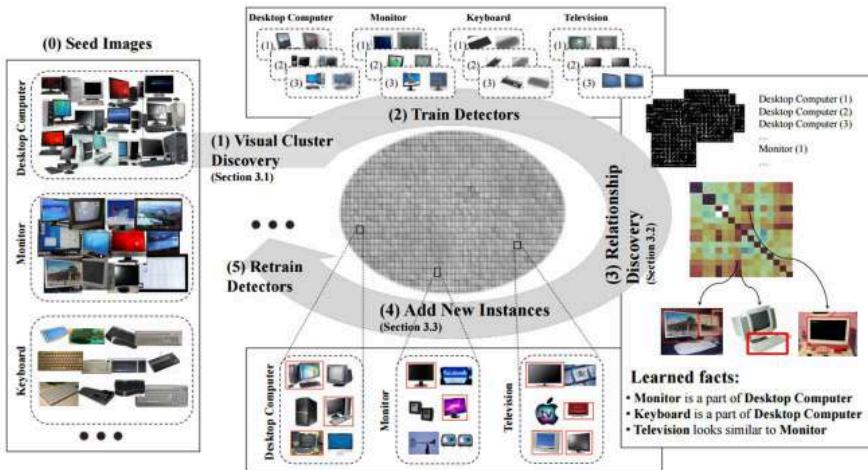
Abbildung 10: Anwendungsbeispiel Recognition with Segmentation: TextonBoost-Verfahren zur automatisierten Detektion von Bildobjekten



Quelle: Shotton et al. 2009: 2. © 2009 Springer.

Die vorgestellten Verfahren lokalisieren und identifizieren Bildobjekte allerdings bisher weitgehend isoliert von ihrem jeweiligen Bildkontext. Zwar spielt die räumliche Konstellation bei den part-based recognition models auch eine Rolle, hier werden jedoch nur diejenigen Teile in Beziehung zueinander gesetzt, die zu einem Objekt gehören. *Context models* machen sich diese Algorithmen zunutze, um auch einzelne Objekte zu kontextualisieren (Sudderth et al. 2008). Anhand von Trainingsbildern können so etwa bestimmte Relationen zwischen Bildobjekten von Straßenszenen (Autos, Bäume, Häuser) in ein geometrisches Modell übersetzt werden, das ein „Verständnis“ für den Gesamtkontext entwickelt und dann einzelne Elemente ähnlicher Szenen erkennen kann (Abb. 11). Obwohl *context models* die Performance der Object Recognition stark erhöhen (Szelski 2010: 712), können die bisher vorgelegten Verfahren bei ihrer Anwendung nur partiell Erfolg aufweisen.

Abbildung 11: Anwendungsbeispiel Context Modell: Lernprozess zur Erkennung eines Arbeitsplatzes anhand einzelner Bildobjekte



Quelle: Russell et al. 2007: 2.

Zusammenfassend lässt sich festhalten, dass Verfahren der Computer Vision bereits heute vielversprechende Möglichkeiten bieten, vorhandene Pixeldaten auch semantischen Konzepten zuzuordnen. Besonders im Bereich der *Objekterkennung* existieren zuverlässige Systeme, die bereits eine größere Anzahl an Objektarten (z. B. Kronen, Schwerter, Autos) unterscheiden können (Bell et al. 2013; Ommer 2011). In diesen Bereich fällt auch die automatisierte Identifikation von Gesichtern (*Gesichtserkennung*; Phillips, Grother & Micheals 2011), Fotos oder Zeichnungen, wie sie etwa Google in vereinfachter Form anbietet. Auch wenn Kohle (2013: 73) anmerkt, dass es bislang nur gelungen sei, „eine ganze Reihe von sehr bestimmten Objektklassen rechnerisch zu identifizieren“ und man „noch weit von einem universellen Algorithmus entfernt“ sei, ergeben sich Chancen, die fortgeschrittenen Analyseverfahren auch für die kommunikationswissenschaftliche Forschungspraxis zu nutzen.

#### 4. Aufklärungspotenzial bisheriger Ansätze der automatisierten Bildinhaltsanalyse für die Kommunikationswissenschaft

Wie bereits im einleitenden Kapitel deutlich wurde, standen kommunikationswissenschaftliche Fragestellungen – entsprechend der Zielsetzung und fachlichen Verortung der Zugänge – bisher nicht im Fokus der methodischen Innovationen. Dennoch bieten die bisherigen Ansätze für die Kommunikationswissenschaft vielfältige Aufklärungspotenziale; insbesondere können die Einsatzbereiche, die Scharkow (2011: 547) für automatisierte Codierverfahren von Texten zusammenfasst, auch für automatisierte Bildinhaltsanalysen beansprucht werden:

1. Computergestützte Verfahren ermöglichen die Verarbeitung von umfangreichen Bildinformationen (maschinenlesbare Bilder und Dokumente, Onlineschriftenstücke, Online-Kommunikation), wodurch sich Reichweite und Genauigkeit der Inferenzschlüsse erhöhen, da der Standardschätzfehler der eigentlich interessierenden Grundgesamtheit (nicht beobachtbare intendierte Aussage des Autors) kleiner wird, je mehr Bilder bzw. Bildinformationen insgesamt codiert werden.
2. Automatisierte Bildcodierverfahren erlauben es, unterschiedliche Operationalisierungsstrategien anzuwenden oder mit manuellen Verfahren zu kombinieren (im Sinne einer Triangulation der Messung), was zu einer Steigerung der Qualität der Analyse und der Validität der Daten führt.
3. Schließlich lassen sich Ressourcen sparen oder an anderer Stelle einsetzen, da bei computergestützten Bildcodierverfahren die Codierkosten nicht mit der Anzahl der Codierungen steigen. Werden etwa offensichtlich zu codierende manifeste Variablen mithilfe eines automatisierten Verfahrens codiert, lassen sich schwierig oder aufwendig zu codierende, mehrdeutige oder bedeutungstragende Bildinhalte von mehreren Personen manuell codieren.

Um sich daneben der Frage nach den Möglichkeiten und Grenzen einer automatisierten Bildinhaltsanalyse als Methode der Kommunikationswissenschaft auf einer *bildinhaltlichen* Ebene zu nähern, soll im Folgenden das *Modell zur theoretischen Dimensionierung von Bildinhalten* (Geise & Rössler 2012) genutzt werden, das mithilfe einer Mehr-ebenenlogik arbeitet und ein adäquates Instrument anbietet, um Bildinhalte intersubjektiv nachprüfbar inhaltsanalytisch zu dimensionieren. Die bereits eingeführten computergestützten Analyseverfahren und zugehörigen Algorithmen werden hierbei entlang der modellierten Bildinhaltsebenen und mit Rückgriff auf die Unterscheidung in Low-Level- und High-Level-Features eingeordnet (vgl. Kapitel 3) und auf ihr Aufklärungspotenzial für die kommunikationswissenschaftliche Forschungspraxis geprüft (vgl. Tabelle 1 für einen Überblick).

##### 4.1 Oberflächenstruktur des Bildes

Auf der Oberflächenstruktur liegen manifeste Bildinhalte, die auf der Darstellungs-, Objekt- sowie Konfigurationsebene standardisiert erfasst werden können (Geise & Rössler 2012).

Auf der *Darstellungsebene* werden *formale Merkmale des Bildes* (Bildgröße, Seitenverhältnis), der *Gestaltung* (Farbe, Kontraste) sowie der *Darstellung* (Kameraperspektive, Einstellung, visuelle Darstellungsstrategien) codiert. Für deren automatisierte Erfassung können Verfahren zur Extraktion von Low-Level-Features zum Einsatz kommen. Mit diesen lassen sich die Datenqualität bzw. Auflösung, Farbwerte, Helligkeits- sowie Farbverteilungen automatisiert erfassen. Damit ein Computersystem Bildwerte

verarbeiten kann, wird das Bild in eine endliche Menge von Zahlenwerten transformiert. Das Ergebnis ist eine Beschreibung des Bildes als „zweidimensionale, regelmäßige Matrix von Zahlen“ (Burger & Bürge 2005: 10; Abb. 4). Formale Merkmale wie Bildgröße oder Seitenverhältnis lassen sich aus der Breite (Anzahl der Spalten) und der Höhe (Anzahl der Zeilen) dieser Bildmatrix extrahieren (ebd.: 11). Für die computergestützte Analyse von Farb- und Kontrastwerten wird auf Pixelwerte zurückgegriffen, die je nach Bildart (Grauwertbild; RGB-Farbbild) unterschiedliche Werte annehmen. Die Farbwerte einzelner Pixel lassen sich einer entsprechenden Farbkategorie zuordnen; dabei wird auch berechnet, wie viele Pixel pro Klasse vorliegen (ebd.: 39). Für Graustufenbilder entstehen so eindimensionale, für Farbbilder dreidimensionale (entlang der Achsen Rot, Grün und Blau) Farbhistogramme (Abb. 5). Mit ihrer Hilfe können Aussagen zur statistischen Häufigkeit der Grau- bzw. Farbwerte sowie zur Belichtung, dem Kontrast, der Dynamik und Sättigung eines Bildes getroffen werden (ebd.: 41ff.).

Derartige Auswertungsverfahren sind problemlos auf die kommunikationswissenschaftliche Forschungspraxis übertragbar. Sie können vor allem im Bereich der automatisierten Selektion von Analysematerial eine Arbeitersparnis bedeuten. So wäre es denkbar, alle schwarz-weißen, unscharfen oder zu kleinen Bildern automatisch auszusortieren. Auf Grundlage exakter Farbhistogramme lässt sich auch identifizieren, wie oft ein spezifisches Bild im Materialkorporus vorkommt, was sich etwa für Forschungen zu Schlüsselereignissen und visuellen Themenkarrieren in der Medienberichterstattung nutzen lässt (Geise et al. 2015). Zudem bietet sich die Möglichkeit der automatisierten Verortung eines Bildes und seiner verhältnismäßigen Größe auf einer Zeitungsseite, um etwa entsprechende Schlüsse zur Relevanz eines Bildes und seines zugehörigen Artikels zu ziehen. Der Einsatz von Farbhistogrammen kann daneben auch für die automatisierte Analyse von Bewegtbildern fruchtbar gemacht werden. So lassen sich etwa Schnitte im Film automatisiert erkennen, wenn sich innerhalb weniger Frames eine starke Veränderung der Farbhistogramme zeigt (Stommel & Müller 2011: 251).

Auf der *Objektebene* werden sachlich fassbare Abbildungsgegenstände wie Personen, Objekte und Situationen erfasst (Geise & Rössler 2012). Zunächst können hierbei Low-Level-Features, wie Kanten- und Kantenrichtungsverteilung, Konturen, Ecken oder Kurven eingesetzt werden, um Bildobjekte zu segmentieren (Abb. 2; Abb. 3). In einem weiterführenden Schritt kann eine Klassifikation dieser gruppierten Einheiten mithilfe von entsprechenden Klassifikationsverfahren vorgenommen werden (Abb. 10). Am Ende dieses Prozesses, bei dem Verfahren der Computer Vision angewandt werden können, um Low-Level-Features in „high level concepts“ (Lew et al. 2006) zu übersetzen, sollte eine möglichst optimale Beschreibung für die erschlossenen Segmente vorliegen. Dabei werden Algorithmen mithilfe von Bilddatensätzen und auf Grundlage des maschinellen Lernens geschult, bestimmte Objekte im Bild zu detektieren (Abb. 7-11). Ein derartiger Algorithmus ist in der Lage „in einer nicht annotierten Sammlung Objekte zu identifizieren, die [...] eine bestimmte Entität enthalten“ (Flores-Herr et al. 2011: 128).

Vor allem die vorgestellten Methoden aus dem Bereich der Computer Vision können je nach Anforderungen eingesetzt werden, um spezifizierbare Abbildungsgegenstände wie Personen (z. B. die auf dem Spiegel-Cover dargestellten Politiker, vgl. Abb. 12), Gegenstände oder Objekte auf der Objektebene automatisiert vom Computer einordnen zu lassen, und versprechen daher großes Aufklärungspotenzial für die Kommunikationsforschung. Dies kann hilfreich für eine entsprechende Vorselektion sein, bietet aber auch Potenzial für eine einfache Auszählung einzelner Objektgruppen. Allerdings wurde gerade komplexen Verfahren eine größere Fehleranfälligkeit attestiert, sodass vor al-

lem die Größe der Anwendungsfelder, die Variabilität der zu analysierenden Phänomene und der Einsatz der passenden Algorithmen über den erfolgreichen Einsatz entscheiden (Stommel & Müller 2011).

Abbildung 12: links: Cover des Time Magazin (Ausgabe 12/2010),  
rechts: Cover des Spiegel (Ausgabe 11/2014)



Für die Oberflächenstruktur eines Bildes ist auch die „räumliche Grammatik“ visueller Kommunikation (Kosslyn et al. 1995) von Bedeutung, da auch durch die Anordnung von Bildelementen zueinander inhaltliche Aussagen transportiert werden (Geise & Rössler 2012). Daher wird auf der *Konfigurationsebene* die objektive Relation der Abbildungsgegenstände als räumliche Konstellation von Objekten, Situationen und Darstellungen zueinander und deren mögliche Bewegungen erhoben (Wang, Joo, Wang & Zhu 2013). Abbildung 12 illustriert, wie die *Konfigurationsebene* Bedeutung transportieren kann. Für eine automatisierte Erfassung bieten *part-based recognition*- und *context*-Modelle Anwendungspotenziale: Entsprechend trainierte Algorithmen könnten etwa die Größe bestimmter Bildobjekte (z. B. der auf dem Spiegel-Cover dargestellten Politiker) identifizieren und ins Verhältnis zueinander setzen (groß/klein; Über-/Unterordnung); dies kann als Basis für eine weitere Interpretation der aus der Kunstgeschichte bekannten *Bedeutungsperspektive* dienen.

#### 4.2 Binnenstruktur des Bildes

Bei der *Binnenstruktur des Bildes* handelt es sich um eine Art Mesoebene, auf der *Bildtypen*, ihre mit-konstituierenden *Symbole* sowie *visuelle Stereotypen* und *Bildtendenzen* erfasst werden (Geise & Rössler 2012). Wie oben bereits angesprochen, können Bilder im Materialkorpus zunächst automatisiert auf die Anwesenheit bestimmter symboli-

scher Objekte (z. B. Flaggen; wie etwa auf dem Time-Magazine-Cover in Abb. 12a) überprüft werden. *Part-based recognition*- und *context*-Modelle, die auch zur Identifikation der Konfigurationsebene dienen, können die Detektion von Objekten mit der Detektion räumlicher Settings, Szenen oder Bildkontexte kombinieren (Cootes et al. 2001; Russell et al. 2007). Dadurch lassen sich bestimmte *Bildtypen* – etwa Interviewsituationen, Landschaftsaufnahmen oder Massendemonstrationen – automatisiert erkennen. Komplexere automatisierte Verfahren können zudem auf extrahierte Features mit höherem semantischen Gehalt zurückgreifen. Auf dieser Basis können Bildern sogar semantische Konzepte (z. B. Sport, Freizeit, Urlaub), fallweise sogar Tendenzen (romantisch, traurig) zugeordnet werden (ProjectOxford 2016; Szegedy 2014) – auch wenn dies bislang nur innerhalb abgeschlossener Domänen befriedigende Ergebnisse erzielt (Flores-Herr et al. 2011: 125). Auf Basis eines Sets aus Bildern, denen im Vorfeld eine bestimmte Tendenz zugeschrieben wurde, lassen sich derartige Algorithmen nach den Regeln des überwachten Lernens trainieren, um dann übereinstimmenden Bildobjekten semantische Konzepte zuzuordnen. Bei der automatisierten Erkennung von Personen-darstellungen und ihrer Tendenz können hierbei bereits gute Identifikationsquoten erreicht werden (abhängig von der Beleuchtung und dem Aufnahmewinkel der abgebildeten Gesichter) (Joo, Steen und Zhu 2015; Stommel & Müller 2011; Sun et al. 2014).

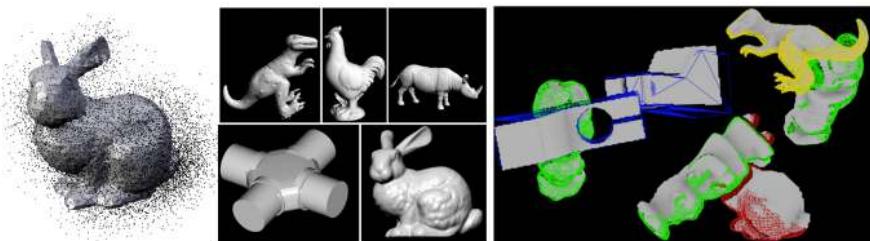
Die Identifikation von Bildobjekten und semantischen Konzepten greift auf grundlegende Ansätze der *instance*, *category* und *part-based recognition* sowie der *recognition with segmentation* zurück, die eigenständig lernen, gemeinsame Features aus einer Grundgesamtheit von Testbildern zu extrahieren, die gemeinsame Objekte bzw. gemeinsame semantische Konzepte beinhalten. Eine solche Identifikation erscheint aber nur plausibel, wenn diese ex ante antizipiert und dem System „beigebracht“ werden kann – „kreative“ Gestaltungen oder neuartige Objektrelationen wie auf dem Cover des Time-Magazins (Abb. 12) dürften damit die Grenzen des Verfahrens markieren. Zudem sind derartig komplexe Verfahren zwar marktreif (Sicherheitsüberwachung; Apps; Militär-anwendungen), besitzen aber noch nicht eine so fortschrittliche Entwicklungsstufe wie Gesichtsidentifikationsalgorithmen und beschränken sich somit bei der Identifikation von Bildobjekten auf berühmte Kunstwerke oder Sehenswürdigkeiten, bekannte Orte, Flaggen, erlernte Logos oder weniger komplexe semantische Konzepte, Stimmungen und Tendenzen (Flores-Herr et al. 2011: 128f.).

Damit kann eine Identifikation von tradierten und sozialisierten stereotypischen Bildinhalten auf *Symbolebene* unter Abhängigkeit von dem jeweiligen extrahierten High-Level-Features und der Bildsammlung befriedigend durchgeführt werden (Ommert 2013), aber eine vollständige Einteilung im Sinne des Dimensionierungsmodells nicht erfüllen. Ähnliches trifft auch auf die Typen- und Tendenzebene zu, die zwar mit annotierten Trainingsdatensätzen und entsprechenden *recognition*- und *context*-Verfahren erfasst werden kann, aber in ihrer Bilddarstellung noch eine zu hohe Variabilität aufweist, um mit den bestehenden Lösungen automatisiert erfasst werden zu können.

#### 4.3 Tiefenstruktur des Bildes

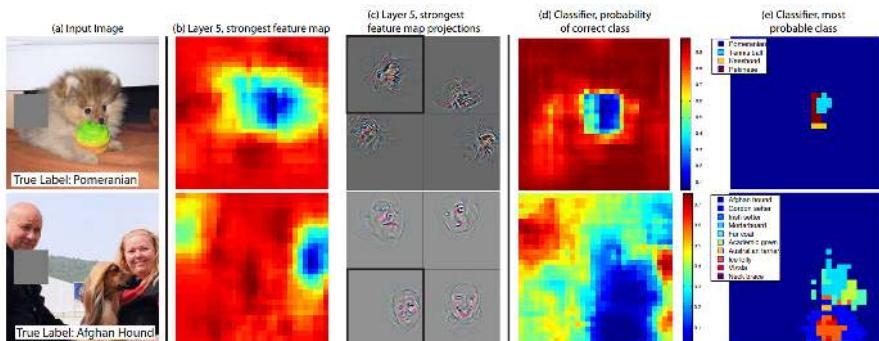
Auf der Ebene der *Tiefenstruktur des Bildes* wird der tiefere Bedeutungssinn des Bild-inhalts in seiner Gesamtheit rekonstruiert (Geise & Rössler 2012). Im Mittelpunkt dieser Bedeutungsschlüsselung steht die Interpretation des Bildinhalts, die als komplexe und oft auch mehrdeutige Sinn- und Wertkonstruktionen bezeichnet und über die dem Rezipienten bestimmte Assoziationen angeboten und Interpretationsmuster nahegelegt werden können. Aufgrund von interpretativen Inferenzen stößt aber bereits der Versuch einer standardisierten *manuellen* Erfassung der Tiefenstruktur an die Grenzen der Me-

Abbildung 13: Anwendungsbeispiel 3D-Modelle: Detektion eines Hasens



Quelle: Drost et al. 2010:7.

Abbildung 14: Anwendungsbeispiel Künstliche neuronale Netzwerke: Identifikation von Bildobjekten



Quelle: Zeiler & Fergus 2014:827.

thodik. Es ist daher wenig verwunderlich, dass damit auch die automatisierte Erfassung der *inhärenten Bedeutungspotenziale* des Bildes enorm begrenzt ist (und vielleicht auch begrenzt bleibt). Dennoch – oder gerade deshalb – ist die weitergehende Bedeutungsentschlüsselung von Bildinformationen Gegenstand aktueller Forschungen.

Das Untersuchungsinteresse richtet sich hierbei insbesondere auf die Fragestellung, wie der *semantic gap* weiter reduziert werden kann (Bell et al. 2013; Chen et al. 2014; Fraunhofer 2015). Aktuelle Weiterentwicklungen im Bereich der Identifikation von Bildobjekten (Armbruster & Hammer 2012; Drost et al. 2010) und -semantiken (Müller & Arens 2010; Münch et al. 2014) arbeiten hierbei etwa mit 3D-Modellen (Abb. 13). In diesem Kontext werden auch sogenannte *künstliche neuronale Netzwerke* entwickelt (Abb. 14), die sich an biologischen Konzepten orientieren und momentan als „State of the Art“ der Computer Vision im Bereich der Klassifizierung und Identifikation von Gesichtern und Objekten gelten (Sun et al. 2014; Szegedy 2014; Zeiler & Fergus 2014).

## 5. Das Bild als Analyseobjekt der automatisierten Inhaltsanalyse – Resümee und zukünftige Herausforderungen

Algorithmenbasierte Bildanalyseverfahren zur Verarbeitung von Bildmatrizen und Pixelwerten, die auf der Grundlage von Basisparametern eines Bildes arbeiten, können auf

der Darstellungsebene zur Anwendung kommen und hier robuste und zuverlässige Ergebnisse liefern (Effelsberg 1999; Smith & Chang 1995; Swain & Ballard 1991; Zabih 1996). Farbhistogramme erlauben präzise Angaben zu Belichtung, Kontrast, Dynamik und Sättigung eines Bildes, die im Vergleich zu einer manuellen Kodierung in höchstem Maße intersubjektiv nachvollziehbar, faktenorientiert und mit geringem Aufwand zu erstellen sind. Auch Aussagen zur Farbverteilung in einem Bild oder zu Filmsequenzen bzw. Schnittgrenzen können automatisiert gewonnen werden. Dies kann die Vorselektion oder -sortierung des Bildkorpus, besonders bei großen Datensätzen, vereinfachen.

Auch für die automatisierte Beschreibung der Objektebene kommt eine Vielzahl computergestützter Verfahren in Frage, die im Zusammenspiel mit großen Bilddatenbanken eine zunehmende Anzahl unterschiedlicher Einzelobjekte zuverlässig erkennen und klassifizieren können (Everingham et al. 2014). Besonders *Gesichtsdetektions-Algorithmen* basieren auf mehrstufigen, selbstlernenden Verfahren und liefern dadurch robuste und zuverlässige Ergebnisse (Degtyarev & Seredin 2010; Stommel & Müller 2011; Szeliski 2010).

Über eine automatisierte Objektdetektion können also bereits einige relevante Bildkategorien erkannt werden, sodass sich ein Einsatz gerade unter den Aspekten der Reliabilität, Validität und Ressourceneinsparung anbietet. Analog zur AIA von Texten lässt sich die ABIA besonders dann verwenden, wenn die Forschungsfrage durch die Auswertung der *Häufigkeit* bestimmter Personen und (einfach zu klassifizierender) Objekte sowie deren Tendenz beantwortet werden kann (Frequenzanalyse). Da auch die standardisierte Bildinhaltsanalyse allgemein immer noch häufig eingesetzt wird, um das Bild auf der Ebene der inhaltlichen Merkmale über relativ grobe, abstrakte thematische Kategorien bzw. Sujet-Klassifikationen zu erfassen (z. B. abgebildete Personen-/gruppen; Geise & Rössler 2012; Lobinger 2012), sollten sich – insbesondere im Kontext eher textfokussierter Inhaltsanalysen – zahlreiche Einsatzmöglichkeiten der ABIA ergeben. Trotz dieser insgesamt vielversprechenden Tendenzen ist die erfolgreiche Detektion von Objekten stark von der Größe der Anwendungsfelder und der Variabilität der zu analysierenden Phänomenologien abhängig.

Über derartige Fragestellungen hinaus stößt die ABIA allerdings – ebenso wie die manuelle BIA – schnell an ihre Grenzen: Gerade weil die notwendigen Algorithmen „Vergleichbares wie die visuelle Wahrnehmung“ leisten können müssten, bei der ein Inferenzschluss auf die gesuchte Bedeutung oftmals eindeutig ist (Tönnies 2005: 15), erweist sich ihre Übertragung in die kommunikationswissenschaftliche Forschungspraxis als äußerst komplex, da sich der Fokus von Bildinhaltsanalysen neben formalen Merkmalen vor allem auf *semantische Konzepte* konzentriert, die für den Computer aus den vorhandenen Pixeldaten nur schwer zu erkennen bzw. zu interpretieren sind.

Nach bisherigem Stand können automatisierte Bildinhaltsanalysen das Bild also insbesondere auf der *Oberflächenstruktur* erfassen (Geise & Rössler 2012); im Ansatz lassen sich auch bestimmte klar definierbare Objekte automatisiert identifizieren, die von ihrem Bedeutungspotenzial her auf der *Binnenstruktur* liegen (insbesondere auf der *Symbolebene*, z. B. Krone, Nationalflagge; ebd.). Anders als die Algorithmen der Gesichtsidentifikation sind Verfahren zur Bestimmung von Bildobjekten und semantischen Konzepten aber noch nicht allgemeingültig genug, um auf breiter Basis auch für die kommunikationswissenschaftliche Forschungspraxis relevante Konzepte zu erkennen. Zwar können den Algorithmen berühmte Kunstwerke, Sehenswürdigkeiten, bekannte Orte, Logos oder bestimmte Stimmungen antrainiert werden, dennoch schränkt dies ihre Anwendbarkeit bei einer kommunikationswissenschaftlich motivierten Bildinhaltsanalyse deutlich ein – mit steigender Komplexität der Inputdaten und Fragestellungen wächst

auch die Anforderung an die Maschine, einen dem Menschen ähnlichen Identifikationsprozess zu emulieren.

Werden Algorithmen jedoch in einem stark abgesteckten Phänomenologiebereich, wie bei der Erkennung von pornografischen Darstellungen (Ni & Huang 2013; Sharma & Pathak 2012), Bildern von Politikern (Joo, Li, Steen & Zhu 2014; Joo, Steen & Zhu 2015) oder Gewalt (Müller & Arens 2010; Münch et al. 2014), eingesetzt, steigen die Chancen auf eine erfolgreiche Identifikation, da hier die Variabilität der semantischen Konzepte deutlich reduziert ist und sie dem Computer mit beschriebenen Lernprozessen beigebracht werden können, wie das Beispiel der mittelalterlichen Schwurfunktionen in historischen Bildern belegt (Bell et al. 2013; Ommer 2011). Den tatsächlichen *Inhalt* des Bildes und seine visuelle Darstellung können die automatisierten Systeme bisher aber (noch) nicht berücksichtigen, noch weniger die weiterführenden Bedeutungspotenziale des Bildes, die sich auf und über die *Tiefenstruktur* des Bildes konstituieren und deren Entschlüsselung auf einer interpretativen Inferenz des Betrachters/Codierers basieren (Geise & Rössler 2012; vgl. Abb. 12a, 12b).

Chancen für die Weiterentwicklung sehen wir in der Idee, die abgebildeten Objekte in ihre wesentlichen Bestandteile zu zerlegen, diese in ihrer Kombinationswahrscheinlichkeit zu klassifizieren und darüber automatisiert zu identifizieren. Zahlreiche Objekte und Objektgruppen lassen sich aus wenigen universalen Grundelementen (z. B. Striche, Punkte, einfache geometrische Formen) zusammensetzen; und gerade Algorithmen, die aufgrund der *kompositionellen Regelhaftigkeit* visuell wahrnehmbarer Information operieren, erscheinen daher vielversprechend. Hierbei würde der Algorithmus zunächst Einzelemente identifizieren, die Lage der Einzelbausteine zueinander klären und zu Bildteilen gruppieren, die zum selben Objekt passen, während er zugleich eine Identifikation für das gesamte Objekt vorschlägt. Die Besonderheit der „räumlichen Grammatik“ bildlicher Kommunikation (Paivio 1991; Kosslyn et al. 1995), durch die Bilder – aufgrund ihrer analogen, räumlich-assoziativen Logik – Bedeutungen nicht nur über die abgebildeten Bildelemente, sondern auch über die jeweilige räumliche Anordnung der Bildelemente zueinander transportieren, erschwert aber auch hier die Automatisierung der Erfassung. Dies gilt umso mehr, wenn sich das Bildmotiv gerade über die Relation einzelner Teilobjekte zu einer Gesamtform zusammensetzt (Ommer 2013). Derartige *emergente Phänomene* gelten gegenwärtig als größte Herausforderung, denn die „manifesten“ Einzelteile enthalten zu wenige und oft widersprüchliche Informationen über das für den Computer „latente“ große Ganze – ein holistischer Ansatz wiederum, mit dem man direkt nach dem Gesamtobjekt sucht, liefert keine verlässlichen Ergebnisse, weil derartige Objekte zu komplex sind und sich zu häufig verändern.

Aus kommunikationswissenschaftlicher Perspektive sehen wir insbesondere auch in der *Rekonstruktion und Definition spezifischer Bildtypen* einen Schlüssel zur Weiterentwicklung der automatisierten Analysen von Medienbildern: Als Bildtypen gelten „spezifische Darstellungsformen, über die Themen, Vorstellungen oder Ereignisse visualisiert werden“ (Grittman 2007: 20) und die von Forschern intersubjektiv nachvollziehbar als Typen klassifiziert werden können (Glassman & Kenney 1994; Griffin 2004; Grittman 2007). Als wiederkehrende, eben typische Darstellungsweisen bestimmter Themen, Aussagen, Situationen oder Handlungsfolgen sollten sich gerade Bildtypen in ihre „manifesten“ Einzelteile zerlegen und automatisiert ihrer „latenten“ Gesamtkonfiguration zuordnen lassen. Werden Bildtypen als Basis der visuellen Mustererkennung eingespeist, würden – im Konzept des Bildtypus quasi inkorporiert (Grittman 2007, 2001; Grittman & Ammann 2011) – neben quantitativ-klassifizierenden Kategorien auch qualitative Aspekte in die Automatisierung integriert. Zwar würden Details des

einzelnen Motivs, Bildelemente und deren innerbildliche Beziehungen zueinander (Grittman 2007: 295) auf dieser automatisierten Analysestufe in den Hintergrund rücken. Doch erlaubte die Abstraktion und Generalisierung auf Bildebene, die im Rahmen der Bildberichterstattung eingesetzten Visualisierungsstrategien bzw. die typischen Merkmale und Strukturen der Medienbilder insgesamt zu erfassen.

Zusammenfassend lässt sich damit ein gegenwärtig noch begrenztes, jedoch stark expandierendes Aufklärungspotenzial bisheriger Ansätze der automatisierten Bildinhaltsanalyse konstatieren. Aufgrund der schnell fortschreitenden technischen Entwicklung auf diesem Gebiet zeigt sich bereits heute, dass neue Algorithmen und Modelle eine immer größere Robustheit gegenüber Störungen aufweisen und immer komplexer werdende Klassen von Bildmustern und semantischen Konzepten automatisiert trainiert und erkannt werden können. Zwar kann die computergestützte Bildinhaltsanalyse die manuelle Codierung im aktuellen Zustand nur ergänzen und dabei für Entlastung im Forschungsprozess sorgen; dennoch führt die rasante Entwicklung dazu, dass die Visuelle Kommunikationsforschung – egal ob zur Vorselektion von Bildmaterial, zur Auszählung einzelner Merkmale oder zur vollständigen Bildinhaltsanalyse – in naher Zukunft von den vorgestellten automatisierten Verfahren profitieren wird.

## Literatur

- Armbruster, W. & Hammer, M. (2012). *Segmentation, Classification, and Pose Estimation of Maritime Targets in Flash-Ladar Imagery*. <http://publica.fraunhofer.de/eprints?urn:nbn:de:0011-n-2196380.pdf> [28.1.2016].
- Ballard, D. H. & Brown, C. M (1982). *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall.
- Bastan, M. & Duygulu, P. (2006). *Recognizing Objects and Scenes in News Videos*. [https://www.researchgate.net/profile/Pinar\\_Duygulu/publication/221369021\\_Recognizing\\_Objects\\_and\\_Scenes\\_in\\_News\\_Videos/links/0deec52e94d582fead00000?origin=publication\\_detail](https://www.researchgate.net/profile/Pinar_Duygulu/publication/221369021_Recognizing_Objects_and_Scenes_in_News_Videos/links/0deec52e94d582fead00000?origin=publication_detail) [28.1.2016].
- Bell, P. (2010). Content Analysis of Visual Images. In: T. van Leeuwen & C. Jewitt (Hrsg.), *Handbook of Visual Analysis* (S. 10-34). Los Angeles/London: Sage.
- Bell, P., Schlecht, J. & Ommer, B. (2013). Nonverbal Communication in Medieval Illustrations Revisited by Computer Vision and Art History. *Visual History* 29(1-2), 26-37.
- Berg, T., Berg, A. C., Edwards, J., Maire, M., White, R., Teh, Y. W., Learned-Miller, E. & Forsyth, D. (2004). *Faces and Names in the News*. [http://acberg.com/papers/berg\\_names\\_and\\_faces.pdf](http://acberg.com/papers/berg_names_and_faces.pdf) [28.1.2016].
- BMBF (2014). *Bewilligte Projekte aus der Bekanntmachung „Mustererkennung“*. <http://www.sifo.de/de/bewilligte-projekte-aus-der-bekanntmachung-mustererkennung-1771.html> [28.1.2016].
- Bock, A., Isermann, H. & Knieper, T. (2011). Quantitative Content Analysis of the Visual. In: E. Margolis & L. Pauwels (Hrsg.), *The SAGE Handbook of Visual Research Methods* (S. 256-282). London: Sage.
- Bohannon, J. (2015). *Facebook Will Soon be Able to ID You in Any Photo*. <http://news.sciencemag.org/social-sciences/2015/02/facebook-will-soon-be-able-id-you-any-photo> [28.1.2016].
- Boykov, Y. & Funka-Lea, G. (2006). Graph Cuts and Efficient N-D Image Segmentation. *International Journal of Computer Vision*, 70(2): 109-131.
- Burger, W. & Bürgel, M. (2005). *Digitale Bildverarbeitung. Eine Einführung mit Java und ImageJ*. Berlin: Springer.
- Chang, S. F. (2002). The Holy Grail of Content-Based Media Analysis. *IEEE Multimedia*, 9(2), 6-10.
- Chen, J., Ye, G., Dong, L. & Chang, S.-F. (2014). *Event-Driven Semantic Concept Discovery by Exploiting Weakly Tagged Internet Images*. [http://www.ee.columbia.edu/ln/dvmm/publications/14/ICMR14\\_FlickrConcept.pdf](http://www.ee.columbia.edu/ln/dvmm/publications/14/ICMR14_FlickrConcept.pdf) [28.1.2016].

- Cootes, T., Edwards, G. J. & Taylor, C. J. (2001). Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine*, 23(6), 681–685, [http://perso.telecom-paristech.fr/~bloch/P6Image/GREC/PAMI01\\_Cootes.pdf](http://perso.telecom-paristech.fr/~bloch/P6Image/GREC/PAMI01_Cootes.pdf) [28.1.2016].
- Datta, R., Li, J. & Wang, J. Z. (2005). *Content-Based Image Retrieval: Approaches and Trends of the New Age*. <http://infolab.stanford.edu/~wangz/project/imsearch/review/ACM05/datta.pdf> [28.1.2016].
- Degtyarev, N. & Seredin, O. (2010). *Comparative Testing of Face Detection Algorithms*. <http://lda.tsu.tula.ru/papers/degtyarev-2010-icisp-ctfd.pdf> [28.1.2016].
- Drost, B., Ulrich, M., Navab, N. & Ilic, S. (2010). *Match Globally, Match Locally: Efficient and Robust 3D Object Recognition*. <http://ar.in.tum.de/pub/drost2010CVPR/drost2010CVPR.pdf> [28.1.2016].
- Effelsberg, W. (1999). Automatische Inhaltsanalyse von digitalen Videos. In: M. Berghaus (Hrsg.), *Interaktive Medien – interdisziplinär vernetzt* (S. 167–193). Wiesbaden: VS Verlag.
- Everingham, M., Eslami, S. M., Van Gool, L., Williams, C. K. I., Winn, J. & Zisserman, A. (2014). *The PASCAL Visual Object Classes Challenge – a Retrospective*. <http://arkitus.com/files/ijcv-14-everingham-pascal.pdf> [28.1.2016].
- Ferrari, V., Tuytelaars, T. & Van Gool, L. (2006). Simultaneous Object Recognition and Segmentation from Single or Multiple Model Views. *International Journal of Computer Vision*, 67(2), 159–188.
- Flores-Herr, N., Sack, H. & Bossert, K. (2011). Suche in Multimediaarchiven von Kultureinrichtungen. In: D. Lewandowski (Hrsg.), *Handbuch Internet-Suchmaschinen* (S. 113–140). Heidelberg: Akad. Verl.-Ges.
- Franz, M. (2007). *Mustererkennung und Klassifikation*. Vorlesung an der Hochschule Konstanz an der Fakultät Informatik und Technische Informatik. [http://www-home.fh-konstanz.de/~mfranz/muk\\_files/lect00\\_Einleitung.pdf](http://www-home.fh-konstanz.de/~mfranz/muk_files/lect00_Einleitung.pdf) [28.1.2016].
- Fraunhofer (2015). *Inhaltsbezogene Videoanalyse*. <http://www.iosb.fraunhofer.de/servlet/is/16900/> [28.1.2016].
- Geise, S. & Rössler, P. (2012). Visuelle Inhaltsanalyse: ein Vorschlag zur theoretischen Dimensionierung der Erfassung von Bildinhalten. *Medien & Kommunikationswissenschaft*, 60(3), 341–361.
- Geise, S. & Rössler, P. (2013a). Die Methode der standardisierten Inhaltsanalyse in ihrer Anwendung auf Bilder. In: T. K. Naab, D. Schütz, W. Möhring & J. Matthes (Hrsg.), *Probleme der Standardisierung und Flexibilisierung – Forschungsprozesse in der Kommunikations- und Publizistikwissenschaft* (S. 80–104). Köln: von Halem.
- Geise, S. & Rössler, P. (2013b). Bildinhaltsanalyse: Die Inhaltsanalyse visueller Inhalte. In: W. Möhring & D. Schütz (Hrsg.), *Handbuch standardisierte Erhebungsmethoden der Kommunikationswissenschaft* (S. 307–328). Wiesbaden: VS-Verlag.
- Geise, S., Weber, N., Siemon, M. & Wischmeyer, K. (2015). Vom Nachrichtenbild zu Medienbild- und Medienmotivikone: Ein Vorschlag zur theoretischen Modellierung des Ikonisierungsprozesses am Beispiel ereignisspezifischer Medienikonen. Vortrag auf der Tagung der DGPuK-Fachgruppe Visuelle Kommunikation, Universität Erfurt.
- Glassman, C. & Kenney, K. (1994). Myths and Presidential Campaign Photographs. *News Photographer*, 49(10), VCQ4-VCQ7.
- Griffin, M. (2004). Picturing America's War on Terrorism in Afghanistan and Iraq. Photographic Motifs as News Frames. *Journalism*, 5(4), 381–402.
- Grittmann, E. & Ammann, I. (2011). Quantitative Bildtypenanalyse. In: T. Petersen & C. Schwenner (Hrsg.), *Die Entschlüsselung der Bilder. Methoden zur Erforschung visueller Kommunikation* (S. 163–178). Köln: von Halem.
- Grittmann, E. & Lobinger, K. (2011). Quantitative Bildinhaltsanalyse. In: T. Petersen & C. Schwenner (Hrsg.), *Die Entschlüsselung der Bilder. Methoden zur Erforschung visueller Kommunikation* (S. 145–162). Köln: von Halem.
- Grittmann, E. (2001). Fotojournalismus und Ikonographie. Zur Inhaltsanalyse von Pressefotos. In: W. Wirth (Hrsg.), *Inhaltsanalyse: Perspektiven, Probleme, Potentiale* (S. 262–279). Köln: von Halem.
- Grittmann, E. (2007). *Das politische Bild*. Köln: von Halem.

- Günther, E., Quandt, T. (2015). Word Counts and Topic Models: Automated Text Analysis Methods for Digital Journalism Research. *Digital Journalism*, 4(1), 75-88.
- Hüftle, M. (2006). *Methoden zur Klassifikation*. <http://www.optiv.de/Methoden/KlassMet/KlassMet.pdf> [28.1.2016].
- Isard, M. & Blake, A. (1998). CONDENSATION – conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1), 5-28.
- Joo, J., Li, W., Steen, F. F. & Zhu, S.-C. (2014). *Visual Persuasion: Inferring Communicative Intent of Images*. [http://cs.ucla.edu/~joo/CVPR14\\_Joo\\_VisualPersuasion.pdf](http://cs.ucla.edu/~joo/CVPR14_Joo_VisualPersuasion.pdf) [28.1.2016].
- Joo, J., Steen, F. F. & Zhu, S.-C. (2015). *Automated Facial Trait Judgment and Election Outcome Prediction: Social Dimensions of Face*. [http://web.cs.ucla.edu/~joo/joo\\_iccv15\\_face\\_election.pdf](http://web.cs.ucla.edu/~joo/joo_iccv15_face_election.pdf) [28.1.2016]
- Keyling, T. (2014). Automatisierte Inhaltsanalyse. In: M. Welker, M. Taddicken, J.-H. Schmidt & N. Jackob, *Handbuch Online-Forschung. Sozialwissenschaftliche Datengewinnung und -auswertung in digitalen Netzen* (S. 233-254). Köln: von Halem.
- Klingemann, H. D. (1984). *Computerunterstützte Inhaltsanalyse in der empirischen Sozialforschung* (Vol. 4). Frankfurt: Campus.
- Kohle, H. (2013). *Digitale Bildwissenschaft*. Glückstadt: Hülsbusch.
- Kosslyn, S., Osherson, D. N. & Gleitman, L. (1995). *Visual Cognition. An Invitation to Cognitive Science*. Cambridge: Massachusetts Institute of Technology.
- Lefebvre, G., Zheng, H. & Laurent, C. (2006). *Objectionable Image Detection by ASSOM Competition*. <http://sites.google.com/site/gregoirelefebvre2/CIVR2006-1.pdf> [28.1.2016].
- Lew, M. S., Sebe, N., Djeraba, C. & Jain, R. (2006). *Content-based Multimedia Information Retrieval: State of the Art and Challenges*. [http://www.ugmode.com/prior\\_art/lew2006cbm.pdf](http://www.ugmode.com/prior_art/lew2006cbm.pdf) [28.1.2016].
- Lobinger, K. (2012). *Visuelle Kommunikationsforschung*. Wiesbaden: VS Verlag.
- Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91-110.
- Marr, D. (1983). *Vision*. San Francisco: Henry & Company.
- Matthews, I., Xiao, J. & Baker, S. (2007). *2D vs. 3D Deformable Face Models: Representational Power, Construction, and Real-Time Fitting*. [http://research.microsoft.com/pubs/69462/2d\\_vs3d\\_ijcv\\_07.pdf](http://research.microsoft.com/pubs/69462/2d_vs3d_ijcv_07.pdf) [28.01.2016].
- Mitchell, W. J. (1986). *Iconology. Image, Text, Ideology*. Chicago/London: University of Chicago Press.
- Müller, M. & Geise, S. (2015). *Grundlagen der visuellen Kommunikation: Theorieansätze und Analysemethoden*. Konstanz: UVK.
- Müller, J. & Arens, M. (2010). *Human Pose Estimation with Implicit Shape Models*. [http://www.researchgate.net/profile/Michael\\_Arens/publication/228975926\\_Human\\_pose\\_estimation\\_with\\_implicit\\_shape\\_models/links/09e415095f8247557000000.pdf](http://www.researchgate.net/profile/Michael_Arens/publication/228975926_Human_pose_estimation_with_implicit_shape_models/links/09e415095f8247557000000.pdf) [28.01.2016].
- Münch, D., Hübner, W. & Arens, M. (2014). *Generalized Hough Transform Based Time Invariant Action Recognition with 3D Pose Information*. [http://akme-a2.iosb.fraunhofer.de/EatThisGoogleScholar/d/2014\\_Generalized%20Hough%20transform%20based%20time%20invariant%20action%20recognition%20with%203D%20pose%20information.pdf](http://akme-a2.iosb.fraunhofer.de/EatThisGoogleScholar/d/2014_Generalized%20Hough%20transform%20based%20time%20invariant%20action%20recognition%20with%203D%20pose%20information.pdf) [28.1.2016].
- Naab, T. & Sehl, A. (2014). Inhaltsanalytische Untersuchungen von User-Generated-Content-Angeboten: Eine Bestandsaufnahme zur Anwendung der Methode. In: K. Sommer, M. Wettstein & W. Wirth (Hrsg.), *Automatisierung in der Inhaltsanalyse* (S. 127-144). Köln: von Halem.
- Ni, P. & Huang, J. (2013). Pornographic Image Filtering Method Based on Human Key Parts. Proceedings of the 2012 International Conference on Information Technology and Software Engineering. *Lecture Notes in Electrical Engineering* 212, 677-688.
- Ommer, B. (2011). *Vom Pixel zum Bild*. <https://www.uni-heidelberg.de/presse/ruca/2011-2/7om.html> [28.1.2016].
- Ommer, B. (2013). *Bilderkennung: Von sehenden Computern*. [http://www.uni-heidelberg.de/exzel\\_lenzinitiative/zukunftskonzept/exchange\\_ommer.html](http://www.uni-heidelberg.de/exzel_lenzinitiative/zukunftskonzept/exchange_ommer.html) [28.1.2016].
- Ommer, B. & Buhmann, J. M. (2007). *Learning the Compositional Nature of Visual Objects*. [http://hci.iwr.uni-heidelberg.de/Staff/bommer/papers/ommer\\_cvpr07.pdf](http://hci.iwr.uni-heidelberg.de/Staff/bommer/papers/ommer_cvpr07.pdf) [28.1.2016].

- Ozkan, D. & Duygulu, P. (2006). *Finding People Frequently Appearing in News*. [http://pdf.aminer.org/000/098/400/finding\\_people\\_frequently\\_appearing\\_in\\_news.pdf](http://pdf.aminer.org/000/098/400/finding_people_frequently_appearing_in_news.pdf) [28.1.2016].
- Paivio, A. (1991). *Images in Mind*. New York: Harvester Wheatsheaf.
- Parks, M. R. (2014). Special Issue: Big Data in Communication Research. *Journal of Communication*, 64, 193–360.
- PennyStocks Lab (2015). *The Internet in Real Time*. <http://pennystocks.la/internet-in-real-time> [28.1.2016].
- Pernkopf, F. (2004). *Klassifikation*. [https://www.spsc.tugraz.at/system/files/specomm04\\_2.pdf](https://www.spsc.tugraz.at/system/files/specomm04_2.pdf) [28.1.2016].
- Phillips, P. J., Grother, P. & Micheals, R. (2011). *Evaluation Methods in Face Recognition*. London: Springer.
- ProjectOxford (2015). Face Detection. <https://www.projectoxford.ai/demo/face#detection> [28.1.2016].
- ProjectOxford (2016). Emotion Recognition. <https://www.projectoxford.ai/demo/Emotion> [28.1.2016].
- Rauchenzauner, E. (2008). *Schlüsselereignisse in der Medienberichterstattung*. Wiesbaden: VS.
- Russell, B. C., Torralba, A., Liu, C., Fergus, R. & Freeman, W. T. (2007). Object Recognition by Scene Alignment. *Advances in Neural Information Processing Systems* 20, 1241-1248.
- Salt, B. (2006). *Moving into Pictures: More on Film History, Style and Analysis*. London: Starwood.
- Sarfraz, M. (2014). *Computer Vision and Image Processing in Intelligent Systems and Multimedia Technologies*. Hershey, PA: Information Science Reference.
- Scharkow, M. (2011). Zur Verknüpfung manueller und automatischer Inhaltsanalyse durch maschinelles Lernen. *Medien & Kommunikationswissenschaft*, 59(4), 545-562.
- Scharkow, M. (2012). *Automatische Inhaltsanalyse und maschinelles Lernen*. epubli.
- Scharkow, M. (2013). Thematic Content Analysis Using Supervised Machine Learning: An Empirical Evaluation Using German Online News. *Quality & Quantity*, 47(2), 761-773.
- Sharma, J. & Pathak, V. K. (2012). Automatic Pornographic Detection in Web Pages Based on Images and Text Data Using Support Vector Machine. *Advances in Intelligent and Soft Computing*, 131, 473-483.
- Sharma, N. & Aggarwal, L. M. (2010). Automated Medical Image Segmentation Techniques. *Journal of Medical Physics/Association of Medical Physicists of India*, 35(1), 3.
- Shotton, J., Winn, J., Rother, C. & Criminisi, A. (2009). *TextonBoost for Image Understanding: Multiclass Object Recognition and Segmentation by Jointly Modeling Appearance, Shape and Context*. <http://jamie.shotton.org/work/publications/ijcv07a.pdf> [28.1.2016].
- Smeulders, A. W., Worring, M., Santini, S., Gupta, A. & Jain, R. (2000). Content-Based Image Retrieval at the End of the Early Years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1349-1380.
- Smith, J. R. & Chang, S-F. (1995). *Single Color Extraction and Image Query*. <http://www.ee.columbia.edu/ln/dvmm/publications/95/smith95b.pdf> [28.1.2016].
- Sommer, K., Wettstein, M. & Wirth, W. (Hrsg.) (2014). *Automatisierung in der Inhaltsanalyse*. Köln: von Halem.
- Sonka, M., Hlavac, V. & Boyle, R. (2014). *Image Processing, Analysis, and Machine Vision*. Stamford, CT, USA: Cengage Learning.
- Stommel, M. & Müller, J. (2011). Automatische, computergestützte Bilderkennung. In: T. Petersen & C. Schwender (Hrsg.), *Die Entschlüsselung der Bilder. Methoden zur Erforschung von visueller Kommunikation. Ein Handbuch* (S. 246–263). Köln: von Halem.
- Sudderth, E. B., Torralba, A., Freeman, W. T. & Willsky, A. S. (2008). *Describing Visual Scenes Using Transformed Objects and Parts*. <http://ssg.mit.edu/~esuddert/papers/ijcv07.pdf> [28.1.2016].
- Sun, Y., Wang, X. & Tang, X. (2014). *Deeply Learned Face Representations are Sparse, Selective, and Robust*. <http://arxiv.org/pdf/1412.1265v1.pdf> [28.1.2016].
- Swain, M. J. & Ballard, D. H. (1991). *Color Indexing*. <http://www.cs.utexas.edu/users/dana/Swain1.pdf> [28.1.2016].
- Szegedy, C. (2014). *Building a Deeper Understanding of Images*. <http://googleresearch.blogspot.de/2014/09/building-deeper-understanding-of-images.html> [28.1.2016].

- Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. London: Springer.
- Tönnies, K. (2005). *Grundlagen der Bildverarbeitung*. München/Boston: Pearson Studium.
- Wang, S., Joo, J., Wang, Y. & Zhu, S.-C. (2013). *Weakly Supervised Learning for Attribute Localization in Outdoor Scenes*. [http://www.stat.ucla.edu/~sczhu/papers/Conf\\_2013/Scene\\_attributes\\_cvpr2013.pdf](http://www.stat.ucla.edu/~sczhu/papers/Conf_2013/Scene_attributes_cvpr2013.pdf) [28.1.2016].
- Wettstein, M. (2014): "Best of both worlds": Die halbautomatische Inhaltsanalyse. In: K. Sommer, M. Wettstein & W. Wirth (Hrsg.), *Automatisierung in der Inhaltsanalyse* (S. 16-39). Köln: von Halem.
- Zabih, R. (1996). *Comparing Images Using Color Coherence Vectors*. <http://www.cs.cornell.edu/~rdz/Papers/PZM-MM96.pdf> [28.1.2016].
- Zeiler, M. D. & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. <https://www.cs.nyu.edu/~fergus/papers/zeilerECCV2014.pdf> [01.05.2016].
- Zheng, H., Daoudi M. & Jedynak, B. (2004). *Blocking Adult Images Based on Statistical Skin Detection*. <http://www.cis.jhu.edu/~bruno/ZhengELCVIA2004.pdf> [28.1.2016].
- Zhu, S.-C. & Mumford, D. (2006). *A Stochastic Grammar of Images*. [http://www.stat.ucla.edu/~sczhu/papers/Reprint\\_Grammar.pdf](http://www.stat.ucla.edu/~sczhu/papers/Reprint_Grammar.pdf) [28.1.2016].

## »methodisch und analytisch sehr fundiert«

Hans Helmut Prinzler, hhprinzler.de 7/2015



**Frauenfiguren im  
zeitgenössischen Bollywoodfilm**  
Im Spannungsfeld zwischen Tradition  
und Moderne

Von Dr. Katharina Görzen  
2015, 344 S., brosch., 64,- €  
ISBN 978-3-8487-1507-7  
eISBN 978-3-8452-5548-4  
(*Filmstudien*, Bd. 70)  
[nomos-shop.de/23125](http://nomos-shop.de/23125)



Unser Wissenschaftsprogramm ist auch online verfügbar unter: [www.nomos-elibrary.de](http://www.nomos-elibrary.de)

Portofreie Buch-Bestellungen unter  
[www.nomos-shop.de](http://www.nomos-shop.de)

Alle Preise inkl. Mehrwertsteuer



**Nomos**

*Abbildung 15: Dimensionierung der Bildebenen, anwendbare computergestützte Analyseverfahren und entsprechendes Anwendungspotenzial für die Kommunikationswissenschaft*

Dimensionierung der Bildebenen	Auf Bildebenen anwendbare computergestützte Analyseverfahren (vgl. Abb. 1)	Anwendungspotenzial für die Kommunikationswissenschaft	
		Low-level-features	High-level-features
Darstellungsebene: Erfassung formaler Merkmale der Gestaltung und der Darstellung, lern- und objektabhängige Größen, Kontrast, Farbe, Textur, Darstellungsstrategien	<p><b>2 Segmentationverfahren, ...</b></p> <ul style="list-style-type: none"> <li>Pixel- und Regionsbasierte Verfahren (Abb. 2)</li> <li>Modell- und Objektabasierte Verfahren (Abb. 3)</li> </ul> <p><b>3 a) Image-Processing-Verfahren, ...</b></p> <ul style="list-style-type: none"> <li>Verarbeitung von Bilduntersichten (Abb. 4)</li> <li>Verarbeitung von Rückprojekten (Abb. 5)</li> </ul> <p><b>3 b) Analyseverfahren, ...</b></p> <ul style="list-style-type: none"> <li>Pattern-Recognition-Verfahren (Abb. 6)</li> <li>Computer-Vision-Verfahren (Abb. 7-11)</li> </ul> <p><b>4 Klassifikationsverfahren,</b></p> <ul style="list-style-type: none"> <li>Überwachte Verfahren</li> <li>Unüberwachte Verfahren</li> </ul>	<p>... um Bildstellen zuerst anhand ihrer Eigenschaften in semantische Einheiten einzuteilen.</p> <ul style="list-style-type: none"> <li>✓ Tiefenbild- und Bildinhalte</li> <li>✓ Extraktions- und Bildinhalte</li> </ul> <p>... um formale Merkmale des Bildes zu analysieren:</p> <ul style="list-style-type: none"> <li>✓ Bildgröße, Seitenverhältnis</li> <li>✓ Farbe, Kontrast</li> <li>✓ Farbe, Beleuchtung, Kontrast, Sättigung</li> </ul> <p>... um Bildmerkmale zu lokalisieren und extrahieren.</p> <ul style="list-style-type: none"> <li>✓ Gezielte Aktion, Objektklassifikation</li> <li>✓ Situationstekennung, Kontexterkennung</li> </ul> <p>... um Bildmerkmale Bedeutungsgehalt zu untersuchen.</p> <ul style="list-style-type: none"> <li>✓ Forscher gibt Codieranweisungen für automatische Analyseverfahren</li> </ul> <p>... um zusammenhängende Bildmerkmale zu selektieren und extrahieren.</p> <ul style="list-style-type: none"> <li>✓ Können räumliche Relationen von Bildmerkmalen zusammenstellen und extrahieren.</li> </ul> <p>... um Zusammenhänge von Bildmerkmale zu erkennen.</p> <ul style="list-style-type: none"> <li>✓ Analysierverfahren, die Abbildungsgesetzmäßigkeiten erkennen.</li> </ul>	
Oberflächenstruktur des Bildes	<p><b>5 Konfigurationsebene: Identifikation objektiver Relation der Abbildungsgegenstände-alternativer Konsolidation der Objekte, Situationen vor abgebildeten Personen, Objekte, Situationen</b></p> <p><b>6 Symbolebene: Identifikation und sozialisierter typischer und stereotypischer Bildinhalte (Konzepte auf der Darstellungs-, Objekt-, sowie Konfigurationsebenen), Identifikation von Beziehungen den Abbildungsgegenständen im Bildraum</b></p> <p><b>7 Typenebene: Identifikation spezifischer, traditionell und sozialisierten Bildarten (kommen auf der Darstellungs-, Objekt- sowie Konfigurationsebene liegen)</b></p> <p><b>8 Innenebene des Bildes</b></p> <p><b>9 Tiefeinstuktur des Bildes</b></p>	<p>... um zusammenhängende Bildmerkmale zu selektieren und extrahieren.</p> <ul style="list-style-type: none"> <li>✓ Können räumliche Relationen von Bildmerkmalen zusammenstellen und extrahieren.</li> </ul> <p>... um Bildinhalten auf Grundlage von Bilddatenbanken zu identifizieren</p> <ul style="list-style-type: none"> <li>✓ Entitätenidentifikation, Geschlechts- und Personenidentifikation</li> <li>✓ Erkennen von weniger komplexe semantischen Konzepten</li> <li>➔ In Abhängigkeit von Trainingsdatensätzen und Variabilität der unterschiedlichen Phänomene</li> </ul> <p>... um Bildtypen auf Grundlage von Bilddatenbanken zu identifizieren</p> <ul style="list-style-type: none"> <li>✓ Entitätenidentifikation, Geschlechts- und Personenidentifikation</li> <li>✓ Erkennen von weniger komplexe semantischen Konzepten</li> <li>➔ In Abhängigkeit von Trainingsdatensätzen und Variabilität der unterschiedlichen Phänomene</li> </ul> <p>... um Bildinhalten auf Grundlage von Bilddatenbanken zu identifizieren.</p> <ul style="list-style-type: none"> <li>✓ Entitätenidentifikation, Geschlechts- und Personenidentifikation</li> <li>✓ Erkennen von weniger komplexe semantischen Konzepten</li> <li>➔ In Abhängigkeit von Trainingsdatensätzen und Variabilität der unterschiedlichen Phänomene</li> </ul> <p>... um Bildinhalten auf Grundlage von Bilddatenbanken zu identifizieren.</p> <ul style="list-style-type: none"> <li>✓ Entitätenidentifikation, Geschlechts- und Personenidentifikation</li> <li>✓ Erkennen von weniger komplexe semantischen Konzepten</li> <li>➔ In Abhängigkeit von Trainingsdatensätzen und Variabilität der unterschiedlichen Phänomene</li> </ul> <p>... um Bildinhalten auf Grundlage von Bilddatenbanken zu identifizieren.</p> <ul style="list-style-type: none"> <li>✓ Entitätenidentifikation, Geschlechts- und Personenidentifikation</li> <li>✓ Erkennen von weniger komplexe semantischen Konzepten</li> <li>➔ In Abhängigkeit von Trainingsdatensätzen und Variabilität der unterschiedlichen Phänomene</li> </ul> <p>... um Bildinhalten auf Grundlage von Bilddatenbanken zu identifizieren.</p> <ul style="list-style-type: none"> <li>✓ Entitätenidentifikation, Geschlechts- und Personenidentifikation</li> <li>✓ Erkennen von weniger komplexe semantischen Konzepten</li> <li>➔ In Abhängigkeit von Trainingsdatensätzen und Variabilität der unterschiedlichen Phänomene</li> </ul> <p>... um Bildinhalten auf Grundlage von Bilddatenbanken zu identifizieren.</p> <ul style="list-style-type: none"> <li>✓ Entitätenidentifikation, Geschlechts- und Personenidentifikation</li> <li>✓ Erkennen von weniger komplexe semantischen Konzepten</li> <li>➔ In Abhängigkeit von Trainingsdatensätzen und Variabilität der unterschiedlichen Phänomene</li> </ul> <p>... um Bildinhalten auf Grundlage von Bilddatenbanken zu identifizieren.</p> <ul style="list-style-type: none"> <li>✓ Entitätenidentifikation, Geschlechts- und Personenidentifikation</li> <li>✓ Erkennen von weniger komplexe semantischen Konzepten</li> <li>➔ In Abhängigkeit von Trainingsdatensätzen und Variabilität der unterschiedlichen Phänomene</li> </ul>	<p>... um Bildstellen zuerst anhand ihrer Eigenschaften in semantische Einheiten einzuteilen.</p> <ul style="list-style-type: none"> <li>✓ Tiefenbild- und Bildinhalte</li> <li>✓ Extraktions- und Bildinhalte</li> </ul> <p>... um formale Merkmale des Bildes zu analysieren:</p> <ul style="list-style-type: none"> <li>✓ Bildgröße, Seitenverhältnis</li> <li>✓ Farbe, Kontrast</li> <li>✓ Farbe, Beleuchtung, Kontrast, Sättigung</li> </ul> <p>... um Bildmerkmale zu lokalisieren und extrahieren.</p> <ul style="list-style-type: none"> <li>✓ Gezielte Aktion, Objektklassifikation</li> <li>✓ Situationstekennung, Kontexterkennung</li> </ul> <p>... um Bildmerkmale Bedeutungsgehalt zu untersuchen.</p> <ul style="list-style-type: none"> <li>✓ Forscher gibt Codieranweisungen für automatische Analyseverfahren</li> </ul> <p>... um zusammenhängende Bildmerkmale zu erkennen.</p> <ul style="list-style-type: none"> <li>✓ Analysierverfahren, die Abbildungsgesetzmäßigkeiten erkennen.</li> </ul>